

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 476

SPRACHANALYSE TERRORISTISCHER GRUPPEN IN SOZIALEN
NETZWERKEN
KORPUSLINGUISTISCHE BETRACHTUNGEN IM KONTEXT
EXTREMISTISCHER AKTIVITÄTEN IM INTERNET

VON
MAXIMILIAN PAUS

SPRACHANALYSE TERRORISTISCHER GRUPPEN IN SOZIALEN
NETZWERKEN
KORPUSLINGUISTISCHE BETRACHTUNGEN IM KONTEXT
EXTREMISTISCHER AKTIVITÄTEN IM INTERNET

VON
MAXIMILIAN PAUS

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Vivien Petras
Humboldt-Universität zu Berlin

Heft 476

Paus, Maximilian

Sprachanalyse terroristischer Gruppen in sozialen Netzwerken:
Korpuslinguistische Betrachtungen im Kontext extremistischer Aktivitäten im
Internet / von Maximilian Paus. - Berlin : Institut für Bibliotheks- und
Informationswissenschaft der Humboldt-Universität zu Berlin, 2021. - 166 S.
: graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und
Informationswissenschaft ; 476)

ISSN 14 38-76 62

Abstract:

Die vorliegende Arbeit widmet sich der Fragestellung, inwieweit sich Personen, die einer bestimmten Ideologie zugeneigt sind, anhand ihres Sprachgebrauches identifizieren lassen. Um dies zu beantworten, werden auf Basis verschiedener korpuslinguistischer Verfahren Indikatoren erarbeitet, die zu einer gesamtumfassenden Analysestrategie führen, um relevante Zielgruppen anhand ihres Sprachgebrauches zu identifizieren, zu skalieren und damit vergleichbar zu machen. Im Rahmen dieser Arbeit werden diese korpuslinguistischen Ansätze auf die Gruppe des Islamischen Staates und deren Sympathisanten in sozialen Netzwerken angewendet und durch ergänzende Techniken wie statistischer Signifikanzanalyse und Sentimentanalyse erweitert. Dafür werden die offiziellen Propagandamagazine des Islamischen Staates, Tweets von deren Sympathisanten und Tweets einer dritten neutralen Gruppe anhand mehrerer Indikatoren auf Gemeinsamkeiten und Unterschiede hin untersucht. Diese Analysen zeigen, dass der Sprachgebrauch der Sympathisanten durchgehend ähnlicher mit dem des Islamischen Staates ist, als dies bei der neutralen Gruppe der Fall ist. Gestützt werden diese Ergebnisse auch durch die Signifikanz- und Sentimentanalysen. Es zeigt sich also, dass der Sprachgebrauch bestimmter Gruppen skalierbar und damit analysierbar ist, was in Bezug auf Counter-Terrorism und wachsende Aktivitäten von Terrorgruppen im Internet neue Möglichkeiten bietet.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im Studiengang Information Science, M. A. an der Humboldt- Universität zu Berlin.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) Lizenz.

Vorwort

Die Idee, eine Masterarbeit zu einem sicherheitspolitischen Thema mit einem eher datenanalytischen Unterbau zu verfassen, kam mir schon zu Zeiten des Bachelorstudiums aus vielerlei Gründen. Seit mehreren Jahren beschäftige ich mich mit sicherheitspolitischen Themen, lese Fachliteratur und habe ein ausgeprägtes Interesse an technischen Aspekten in diesem Zusammenhang entwickelt. Durch die Erfahrungen und Inspirationen, die ich während des Bachelorstudiums an der Heinrich-Heine-Universität in Düsseldorf und besonders während des Masterstudiums am Institut für Bibliotheks- und Informationswissenschaft (IBI) der Humboldt-Universität zu Berlin sammeln konnte, zeichnete sich für mich sehr schnell ab, welche vielseitigen Möglichkeiten Informationswissenschaft in diesem Themenkomplex mit sich bringen kann. Diese Ebenen in Form einer Masterarbeit miteinander in Einklang zu bringen, stellte sich dabei vor allem zu Beginn der Arbeit als schwierig heraus, da sich Probleme zeigten, auf die ich in meiner gesamten Zeit an einer Universität noch nie gestoßen war. Die Lösung dieser Probleme habe ich zu einem großen Teil der lehrstuhlübergreifenden Unterstützung durch Dozenten und Mitarbeiter des gesamten Instituts zu verdanken, denen ich im Folgenden meinen Dank aussprechen möchte.

Zuerst möchte ich meinen Dank Frau Prof. Vivien Petras, PhD und Frau Dr. Maria Gäde aussprechen, die mich während des gesamten langen Prozesses unterstützt und motiviert haben und auch die Begutachtung dieser Arbeit übernehmen. Für die enge, konstruktive und stets freundliche Betreuung während des gesamten Studiums auch außerhalb dieser Arbeit bin ich Ihnen äußerst dankbar und auf immer verbunden und möchte dies herausheben.

Außerdem bedanke ich mich herzlich bei Prof. Dr. Robert Jäschke, der mir mit seiner Datensammlung ermöglicht hat, einen Korpus mit äußerst relevanten Daten zu erstellen, die seit Jahren nicht mehr verfügbar sind. In diesem Zusammenhang bedanke ich mich auch bei Michael Paris, M.Sc. für

die technische Unterstützung und die außerordentlich schnelle Bereitstellung des Zugangs auf diesen Servern.

Zuletzt geht mein Dank an das IBI und all dessen Mitarbeiter und Mitarbeiterinnen, von denen ich unbeschreiblich viel lernen und mit denen ich eine wunderschöne Zeit verbringen durfte. Des Weiteren sind an dieser Stelle Michael Trellenkamp und Lars Kasischke hervorzuheben, die mich bei der Erstellung dieser Arbeit in gleich mehrfacher Hinsicht unterstützt haben und im Rahmen von wertvollen Diskussionen nicht wenige Denkanstöße eingeleitet haben.

Berlin, 10.11.2020

Inhaltsverzeichnis

1	Einleitung	12
1.1	Entstehung der Internetpräsenz des „Islamischen Staates“	16
1.2	Notwendigkeit der Terrorismusbekämpfung im Netz	18
2	„War in Cyberspace“ und Terrorismusbekämpfung im Netz	22
2.1	Forschungsstand im Kontext der Online-Terrorismusbekämpfung	24
2.2	Problemstellungen des Monitorings im Internet	33
2.2.1	Technische Dimension	34
2.2.2	Linguistische Dimension	39
2.3	Online-Strategie des „Islamischen Staates“	44
2.3.1	Propagandamagazine <i>Dabiq</i> und <i>Rumiyah</i>	48
2.3.2	Twitter als Kommunikationsmedium für Sympathisanten	53
2.3.3	Gruppenstruktur und die Rolle der Frauen	60
3	Analyse in Form korpuslinguistischer Sprachanalyse	67
3.1	Ziele und Chancen einer korpuslinguistischen Analyse	68
3.2	Weiterführende Methoden und Strategien zur Ergebnisverifizierung	71
3.3	Datenerhebung und Korpusaufbau	73
3.4	Sprachanalyse und Anwendung der Indikatoren	79
3.5	Weitergehende Analysen	93
4	Diskussion und Fazit	100
4.1	Korpuslinguistische Sprachanalyse	100
4.2	Weitergehende Analysen	104
4.3	Weitergehende Erkenntnisse und zukünftige Möglichkeiten	107
5	Zusammenfassung	111
	Literaturverzeichnis und sonstige Quellen	115
	Anhang	120
	Eidesstattliche Erklärung	167

Abbildungsverzeichnis

Abbildung 1: Informationsfluss auf dem Weg zur Intelligenz	37
Abbildung 2: Twitter storm am Beispiel des „Islamischen Staates“	55
Abbildung 3: Cyclical cognitive reinforcement	61
Abbildung 4: Vergleich der relativen Häufigkeiten von sieben Termen in Promille	83
Abbildung 5: Prozentuale relative Termhäufigkeit verteilt nach Magazinen	85

Tabellenverzeichnis

Tabelle 1: Dokument- und Tokenverteilung der jeweiligen Korpora	80
Tabelle 2: Top-20 Terme (+ absolute Häufigkeit) in allen Korpora	81
Tabelle 3: Top-20 Terme (+ relative Häufigkeit in Promille) in allen Korpora	82
Tabelle 4: Kontingenztabelle beobachteter Werte des Terms „allah“	94
Tabelle 5: Kontingenztabelle erwarteter Werte des Terms „allah“	95
Tabelle 6: Chi-Quadrat-Prüfwerte der Terme in Bezug auf Referenzkorpus A	96

Abkürzungsverzeichnis

AQ	–	Al-Qaida
AQA	–	Al-Qaida in Afghanistan
AQAP	–	Al-Qaida auf der arabischen Halbinsel
AQI	–	Al-Qaida im Irak
CT	–	Counter-Terrorism
IS	–	Islamischer Staat
ISIL	–	Islamischer Staat im Irak und der Levante
LW	–	Lone Wolves
NLP	–	Natural Language Processing
OM	–	Online-Monitoring
RE	–	Regular Expressions
SA	–	Sentimentanalyse
SOCMINT	–	Social Media Intelligence
TG	–	Testgruppe
UGC	–	User-generated-content
ZP	–	Zielperson

1 Einleitung

„*The spark has been lit here in Iraq, and its heat will continue to intensify – by Allah’s permission – until it burns the crusader armies in Dabiq*“ (Dabiq, 2014a, S. 3).

Dieses Zitat von Abu Mus’ab az-Zarqawi, einem der weltweit meistgesuchtesten Terroristen des islamischen Fundamentalismus, zeigt in drastischer Weise die Zielsetzung der Anhänger des sogenannten „Islamischen Staates“ (IS). Diese Gruppe verkündete im Juni 2014 infolge des syrischen Bürgerkrieges die Ausrufung eines Kalifats auf irakischem und syrischem Staatsgrund und ist seit den darauffolgenden Jahren vor allem durch zahlreiche Massaker an Zivilisten und ausländischen Gefangenen, sexuelle und ökonomische Ausbeutung sowie der Zerstörung von Kulturstätten in den weltweiten Fokus geraten (Richards, 2016b, S. 1). Besonderes Augenmerk gilt dabei seit jeher der bis dahin bei Terrorgruppen eher seltenen und ungewöhnlichen Nutzung von sozialen Medien, die bei der Kommunikation zwischen den Akteuren und deren Anhängern des IS eine entscheidende Rolle spielt. So werden über das Internet nicht nur Informationen und Neuigkeiten verbreitet, sondern wird auch aktiv um Rekruten geworben sowie Propagandamaterial verbreitet, sodass sich der „Islamische Staat“ innerhalb kürzester Zeit zum weltweit führenden Lieferanten von jihadistischer Indoktrination entwickeln konnte (Hamblet, 2017, S.1). Diese Rekrutierungsstrategie zeigte bereits insofern Wirkung, als dass sich tausende ausländische Kämpfer dieser Gruppe anschlossen und weltweit Terroranschläge verübten (Richards, 2016b, S. 1).

Besonders wichtig für diese Online-Strategie des IS sind einerseits soziale Netzwerke wie Twitter, andererseits die beiden Propagandamagazine *Dabiq* und *Rumiyah*, deren Namen auf Orte zurückgehen. *Dabiq* ist eine Ortschaft in der Region Aleppo im nördlichen Syrien und spielt in der Ideologie des IS eine besondere Rolle, da dort nach fundamentalistischer Leseart am Ende der

Zeit die muslimischen Armeen auf ihre Feinde treffen werden (Ingram, 2016, S. 466). Ähnliches gilt für *Rumiyah*, das der arabische Name für die italienische Hauptstadt Rom ist und ebenfalls als Ziel der jihadistischen Expansion ausgerufen wurde (Ingram, 2018, S. 21). Diese Menge von verbreitetem Material, das von Kurznachrichten in sozialen Medien über audiovisuelle Medien und Computerspiele bis hin zu eigenen Propagandazeitungen reicht, ist auch im Fokus der Sicherheitsbehörden und Institutionen sowie der Forschung und Wissenschaft, die sich mit immer neuen Ansätzen und Strategien zur Terrorismusbekämpfung (Counterterrorism; im Folgenden CT) beschäftigen. Ein besonders wichtiges Ziel in diesem Zusammenhang ist das Online-Monitoring (OM) von Akteuren, die entweder terroristischen Gruppierungen angehören oder diesen zumindest nahestehen oder auch manchmal alleine handeln, um zukünftige Terroranschläge vereiteln zu können und auf staatsgefährdende Aktionen reagieren zu können.

Dabei bieten sich eine Vielzahl von möglichen Ansätzen und Instrumenten, die auf verschiedensten Datensätzen beruhen und unterschiedlichste Ergebnisse liefern. Eine besonders vielversprechende, aber in der Forschung noch nicht sehr verbreitete Methode ist die Korpusanalyse auf Basis mehrerer Textkorpora, die das Kommunikationsverhalten und den Sprachgebrauch einer bestimmten Zielgruppe analysiert und die im weiteren Verlauf dieser Arbeit genauer betrachtet wird. Die Relevanz dieses Themenkomplexes liegt vor allem darin, als dass auch Jahre nach der militärischen Niederlage des „Islamischen Staates“ im Irak und Syrien von einer Niederlage der jihadistischen Ideologie keine Rede sein kann, wie das weitere Vorhandensein einer Vielzahl von fundamentalistischen Terrorgruppen und deren Internetaktivitäten zeigt. Des Weiteren bietet sich durch die Weiterentwicklung bereits existierender Analyseverfahren und der Kombination ebendieser mit neuen Methoden die Möglichkeit einer

Verbesserung bereits bestehender Sicherheitsstrategien. Jedoch ist die Kombination von Korpusanalysen mit anderen Analyseverfahren bisher ein in der Forschung wenig betrachtetes Themengebiet, obwohl sich gerade diese Methodik aufgrund der Vielzahl von Textmaterial aus Reihen des IS anbietet. Aus diesem Grund widmet sich die vorliegende Arbeit einer Reihe von Forschungsfragen, die sich in den Kontext sicherheitspolitischer Problemstellungen einreihen und mithilfe korpuslinguistischer Methoden beantwortet werden sollen. Von besonderer Relevanz in diesem Zusammenhang ist die Frage, ob sich Personen, die einer bestimmten Ideologie zugeneigt sind, anhand ihres Sprachgebrauchs identifizieren lassen. Außerdem liegt der Fokus auf korpuslinguistischen Verfahren und der Fragestellung, ob sich anhand dieser konkrete Indikatoren erarbeiten lassen, die besonders ergiebig für eine Identifizierung solcher Personen sind. Abschließend wird betrachtet, welche Unterschiede, beziehungsweise Gemeinsamkeiten sich zwischen verschiedenen Testgruppen (TG) feststellen lassen können. Dies ist besonders von Interesse, da sich aufgrund der unterschiedlichen Textkorpora, die auf unterschiedlichen TG beruhen, Rückschlüsse auf den Sprachgebrauch dieser betrachteten Gruppen ziehen lassen. Mit der Beantwortung dieser Fragen lässt sich im Idealfall die wirksame Anwendung von korpuslinguistischen Methoden auf sicherheitsrelevante Themenkomplexe nachweisen. Der Mehrwert besteht darin, dass nicht nur eine auf jihadistische Terroristen zugeschnittene Methode, sondern vielmehr eine Analysestrategie vorgestellt wird, die gruppenunspezifisch ist und die sich auf jegliche Personen oder Gruppen anwenden lässt, die Textmaterialien produzieren. Abschließend werden im Rahmen dieser Arbeit Textkorpora zusammengestellt und verwendet, die sich unter Umständen für weitere Forschungen mit korpuslinguistischen Methoden in diesem Gruppenumfeld weiterverwenden und verbessern lassen. Methodisch sollen diese Fragen beantwortet werden, indem drei Textkorpora

aus unterschiedlichen Datenquellen aufbereitet werden und dabei jeder Textkorpus eine bestimmte Gruppe repräsentiert, die im Folgenden erläutert wird. Der erste Korpus besteht aus Textmaterial, das direkt vom IS in Form der Propagandamagazine *Dabiq* und *Rumiyah* verbreitet worden ist. Der zweite Korpus repräsentiert Sympathisanten und Anhänger dieser Gruppe, die im Internet Textmaterial in Form von Tweets im sozialen Netzwerk Twitter produzieren. Abschließend besteht der dritte Korpus aus Tweets, die sowohl aus muslimischen, als auch nicht-muslimischen Kreisen kommen und sich zu thematisch ähnlichen Sachverhalten auf Twitter äußern. Nach der Aufbereitung dieser Textkorpora soll eine Reihe von korpuslinguistischen Analysen durchgeführt werden, um Unterschiede zwischen den jeweiligen Korpora und deren Nähe zum IS-Referenzkorpus durch Indikatoren herauszustellen, wobei stets die genutzte Sprache im Fokus steht. Diese Ergebnisse sollen im weiteren Verlauf durch Signifikanztests und eine Sentimentanalyse (SA) abgerundet werden, um eventuelle Unterschiede auf mehreren Dimensionen darstellen zu können.

Der Aufbau der vorliegenden Arbeit gliedert sich folgendermaßen: Zunächst steht die Entwicklung der fundamentalistischen Propaganda im Internet im Fokus, um anschließend die daraus folgende Notwendigkeit des CT zu verdeutlichen. Im darauffolgenden Teil der Arbeit geht es thematisch um CT im Internet, wobei zunächst der Forschungsstand und thematische Kontext der Terrorbekämpfung im Netz betrachtet wird, um sich anschließend Problemstellungen in diesem Bereich zu widmen. Anschließend wird die Online-Strategie des IS näher untersucht, wobei die Propagandamagazine und Twitter unabhängig voneinander behandelt werden. In diesem Zusammenhang wird auch der Gruppenstruktur der Akteure sowie der Rolle der Frauen innerhalb der Gruppe Beachtung geschenkt. Im anschließenden Kapitel beginnt der methodische Teil der Arbeit, in dem sich zunächst den Zielen einer korpuslinguistischen Analyse zugewendet wird und dann

weiterführende Methoden zur Ergebnisverifizierung betrachtet werden. Darauf aufbauend wird die Datenerhebung nachvollzogen sowie der Korpusaufbau im Detail erklärt, um dann zu den Analysen zu kommen, die aus korpuslinguistischen, statistischen Sprachanalysen und einer SA bestehen. Abschließend werden die Ergebnisse kritisch betrachtet und darauf aufbauend Vorschläge erarbeitet, wie eine weitere Behandlung dieses Themenbereiches verlaufen könnte.

1.1 Entstehung der Internetpräsenz des „Islamischen Staates“

Das offensive Auftreten des IS im Internet begann nicht etwa erst mit der Ausrufung des Kalifats im Juni 2014, sondern hat seine Anfänge bereits Jahre vorher zu verzeichnen und steht in direktem Bezug zu anderen jihadistischen Terrorgruppen. Eine dieser Gruppen nutzte das Internet bereits weit vorher und ist als al-Qaida (AQ) zu trauriger Bekanntheit gelangt. Diese Gruppe produzierte bereits seit 2004 qualitativ minderwertige Videos von Kampferfolgen sowie Monologen von ihrem Anführer Osama bin Laden, die dann wenig später nicht selten von Internetcafés in Pakistan aus ins Internet hochgeladen wurden, aber trotzdem in westlicher Berichterstattung nicht an Wirkung verfehlten. Die zu dieser Zeit bereits im Jemen wachsende Gruppe al-Qaida auf der arabischen Halbinsel (AQAP) veröffentlichte mit dem elektronischen Magazin *Inspire* eine dazu passende Anleitung für den Do-It-Yourself-Jihad (Veilleux-Lepage, 2016, S. 36). Diese ersten modern verbreiteten Formen von fundamentalistischem Material dienten als Vorbild für die spätere Online-Strategie des IS, wenngleich der Ursprung des Jihad im Internet noch weiter zurückreicht.

So gilt die Seite *Azzam.com* als die allererste authentische Webseite in Form eines Forums mit islamistischen Mitgliedern, die zur Zeit des ersten Tschetschenien-Krieges 1994-1996 entstand und was nicht zuletzt durch die Beschreibung „*independent media organization providing authentic news and*

information about jihad and the Foreign Mujahideen everywhere“ deutlich gemacht wurde (Veilleux-Lepage, 2016, S. 37). Die Entstehung solcher Webseiten führte bis in die 2000er Jahre zu bis zu 4300 aktiven Jihadistenforen im Internet und wurde erst mit der wachsenden Popularität von File-Sharing-Portalen und sozialen Netzwerken unbedeutender. Der Ableger al-Qaida im Irak (AQI), der auch als ein erster Vorläufer des IS gilt, war infolgedessen die erste Gruppe, die Material zum Download anbot und die Möglichkeiten des Web 2.0 erkannte, sowie schockierende Videos von Enthauptungen auf solchen Plattformen veröffentlichte, die millionenfach konsumiert wurden (Veilleux-Lepage, 2016, S. 38f). Infolge des Todes von Osama bin Laden im Mai 2011 wurden diese Aktivitäten noch zusätzlich verstärkt, da einige bis dahin von Terrorgruppen zur Verbreitung des Propagandamaterials installierte Medienzentren das Internet als „Schlachtfeld des Jihad“ ausgerufen haben, was nicht zuletzt dem Ziel diente, sich von Massenmedien wie al-Jazeera, auf deren Berichterstattung man lange angewiesen war, unabhängig zu machen. Diese Medienzentren, die sich zu großen Teilen mit Inhalten aus den Jihadistenforen speisten, waren dann auch die ersten Gruppeninstanzen, die durch die Erstellung von Namen, Logos und eigenen Designs eine Professionalisierung anstrebten und erste Profile im sozialen Netzwerk Twitter eröffneten (Weimann, 2019, S. 27f).

AQI, die sich kurz darauf bereits aufgrund der Vereinigung mit anderen Terrorgruppen in „Islamischer Staat im Irak und der Levante“ (ISIL) umbenannten, haben mit der Erstellung von offiziellen Profilen in sozialen Netzwerken bereits eine eigene Form der Kommunikation mit ihren Anhängern beabsichtigt (Cohen, Kruglanski, Gelfand, Webber, & Gunaratna, 2018, S. 146). Nach der Ausrufung des Kalifats im Juni 2014 und der anschließenden Umbenennung in IS wurde auch die von da an bekannte „*News Agency Amaq*“ gegründet, die als offizielles Sprachrohr des IS gilt und bis heute aktiv ist (Weimann, 2019, S. 29f). In dieser Form konnte sich der IS

zu einer der weltweit einflussreichsten und gefürchtetsten Terrorgruppen entwickeln, was zu großen Teilen auf ihre professionelle und hochentwickelte Kommunikation mit der Außenwelt zurückzuführen ist, auf die sehr einfach zuzugreifen ist, die hochgradig anziehend für Sympathisanten wirkt und die in mehreren Sprachen verfasst ist. Eines der ersten Produkte, die im Rahmen dieser Strategie entstanden, war das Terrormagazin *Dabiq*, auf das im weiteren Verlauf dieser Arbeit noch Bezug genommen wird und welches beabsichtigt, den Leser in fundamentalistischer Leseart zu erziehen und über Pläne und Projekte im Kalifat aufzuklären (Veilleux-Lepage, 2016, S. 41). Ebenso wie die Aktivitäten in sozialen Netzwerken zielen diese Propagandamechanismen des IS darauf ab, Sympathisanten im Um- und Ausland für ihre bewaffneten Kräfte und für das Kalifat zu gewinnen, beziehungsweise im Ausland zu Attacken auf die Zivilbevölkerung anzustiften, wie eine Vielzahl von Terroranschlägen außerhalb des eigenen Territoriums leider zeigen (Richards, 2016b, S. 1).

1.2 Notwendigkeit der Terrorismusbekämpfung im Netz

Wie die bereits skizzierte Miteinbeziehung des Internets als Schauplatz des internationalen Terrorismus zeigt, stellt sich auch im Rahmen des CT die Frage, wie darauf reagiert werden sollte. Durch eine Vielzahl von Veränderungen, die im Rahmen der Globalisierung und der technischen Innovationen der letzten Jahre entstanden sind, hat sich das Umfeld der Sicherheitsinstitutionen massiv verändert, was vor allem durch politische, militärische und ökologische Konflikte zusätzlich verschärft wurde (Șuşnea, 2018, S. 427).

Eine besondere Gefahr geht dabei von Personen aus, die fundamentalistisch oder ideologisch orientiert sind, keiner hierarchischen Gruppenstruktur angehören und autonom terroristische Aktivitäten planen könnten, dabei aber in der virtuellen Unsichtbarkeit Schutz finden. Solche Personen werden auch

als „Lone Wolves“ (LW) bezeichnet und stellen Sicherheitsinstitutionen vor besondere Schwierigkeiten (Striegher, 2013, S. 35). Am Beispiel des IS wird schnell deutlich, dass die Gefahren durch solche LW nicht unterschätzt werden dürfen, wie eine Vielzahl von Terroranschlägen zeigt, die von diesen durchgeführt wurden und dem Ruf des IS folgten – nicht selten nach dem Konsum seiner Online-Propaganda.

Obwohl im Rahmen militärischer Operationen die physische Präsenz des IS in seinem ausgerufenen Kalifat zurückging, besitzt die Terrorgruppe nach wie vor eine starke Online-Präsenz (Macnair, & Frank, 2018a, S. 110). Durch die Nutzung von verschiedenen Kommunikationstechnologien ist es solchen Terrorgruppen sehr einfach, dezentral und von völlig unterschiedlichen Standorten aus ihre Aktivitäten zu planen oder LW zu erreichen, beziehungsweise Sympathisanten zu solchen zu machen (Striegher, 2013, S. 41). Dies geschieht in erster Linie durch die Anziehung potentieller Rekruten in Form ihrer Propaganda, die als Spektakel ausgelegt ist und darauf abzielt, sympathische oder oppositionelle Gefühle beim Rezipient zu erzeugen (Richards, 2016b, S. 1f).

Dabei machen sich solche Terrorgruppen nicht selten Strukturen zunutze, wie sie in den Ländern ihrer Feinde vorzufinden sind. So sind beispielsweise freie Medien in einer offenen Gesellschaft besonders anfällig für die Manipulation und Instrumentalisierung durch Terroristen, die ohne die Aufmerksamkeit dieser Medien gar nicht erst in diesem Maße in der Lage wären, ihre Botschaften in Form ihrer manipulierenden Propaganda zu verbreiten. Dazu gehört ebenfalls, dass im Rahmen von Terroranschlägen häufig weiche Ziele wie der öffentliche Nahverkehr, Einkaufszentren oder Veranstaltungen als mögliches Terrorziel in Betracht gezogen werden, was es den Sicherheitsinstitutionen zusätzlich erschwert, auf solche Gefahren zu reagieren (Striegher, 2013, S. 42f). Durch dieses Zusammenspiel leicht zugänglicher islamistischer Propaganda und psychologischer Sogwirkung für

Sympathisanten sowie nur spärlich bewachten weichen Anschlagzielen entsteht ein sich stets erneuernder Kreislauf von Neurekrutierung und Gefahrenpotential, der nur äußerst schwer zu durchbrechen und überhaupt schwer zu identifizieren ist.

Eine Möglichkeit der Begegnung dieser Gefahren besteht darin, die Verbreitung von Propagandamaterial, das im Internet zu finden ist und dem Ziel dient, zu solchen Anschlägen anzustiften oder Personal für diese Pläne zu gewinnen, bestmöglich zu unterbinden. Eine Schwierigkeit ist jedoch, dass die Organisationsstruktur des IS sehr inkonstant – ähnlich einem Bienenschwarm, der sich ständig neu reorganisiert – ist, und selbst nach einer Löschung eines Kontos in sozialen Netzwerken, das der Propagandaverbreitung dient, selbiges schnell wieder neueröffnet und die Verbreitung weiter stattfindet, sofern denn ein Konto von Behörden überhaupt als solches entdeckt wird. Dieses Verhalten ermöglicht dem IS eine anhaltende Online-Präsenz, die nur schwer zu unterbinden ist (Ashcroft, Fisher, Kaati, Omer, & Prucha, 2015, S. 161).

Es zeigt sich also, dass durchaus Zugriffspunkte existieren, um den Aktivitäten terroristischer Gruppen Einhalt zu gebieten, wenngleich diese Gruppen schwierig zu identifizieren sind. Problemstellungen in diesem Kontext können, wie das Beispiel des IS zeigt, nicht allein militärisch gelöst werden, sondern verlangen nach zusätzlichen Lösungsansätzen, die nicht nur auf das Informationsbedürfnis militärischer Kapazitäten zugeschnitten sein sollten (Şuşnea, 2018, S. 430). Des Weiteren zeigt sich, dass eine Analyse von Inhalten in sozialen Netzwerken zu einer wichtigen Aufgabe von Sicherheitsinstitutionen geworden ist. Dabei liegt eine Schwierigkeit nicht nur in der Menge der produzierten Inhalte, sondern auch in deren Verbreitungsform und Sprachvielfalt (Ashcroft et al., 2015, S. 162). Daher werden Methoden, die im Rahmen von sozialen Netzwerken eingesetzt werden und auch als Social Media Intelligence (SOCMINT) bezeichnet

werden, als entscheidende zukünftige Quelle für Sicherheitsinstitutionen erachtet (Benigni, Joseph, & Carley, 2017, S. 3).

2 „War in Cyberspace“ und Terrorismusbekämpfung im Netz

Mit dem bereits oben skizzierten vermehrten Vorkommen von Aktivitäten terroristischer Gruppen im Internet und der daraus resultierenden Notwendigkeit der Terrorbekämpfung mithilfe technischer Überwachungsmethoden wird dieser Teil der Arbeit dem theoretischen Unterbau des Themenkomplexes gewidmet. Der Fokus liegt dabei zunächst darauf, bereits durchgeführte Forschungen zu betrachten, die methodisch oder inhaltlich verwandt mit der Terrorismusbekämpfung im Internet sind. Dadurch soll einerseits der Kontext, in dem diese Arbeit steht, andererseits aber auch die Vielseitigkeit der Methoden verdeutlicht werden, da viele dieser ihren Ursprung nämlich in inhaltlich völlig anderen Themenbereichen haben. Besonders interessant ist die Tatsache, dass solche Methoden universell für verschiedene Anwendungsgebiete einsetzbar sind.

Nach dem Forschungsstand geht es um Problemstellungen, die gleich auf mehreren Ebenen beim OM auftreten können. Neben technischen Schwierigkeiten, die sich in einem so sensiblen Bereich als folgeschwer erweisen können, gibt es auch eine Vielzahl von Problemen, die aus linguistischer Perspektive betrachtet werden müssen. Die linguistische Dimension stellt Anwender und Entwickler dabei vor besonders große Hürden, da diese – im Gegensatz zu technischen Problemen – häufig nicht oder nur äußerst schwer zu lösen sind.

Anschließend liegt der Fokus auf der Online-Strategie des „Islamischen Staates“, die ebenfalls mehrere Ebenen umfasst. Zunächst wird ein Blick auf die offiziellen Magazine geworfen, die hochprofessionell und mit großer Wirkungsabsicht veröffentlicht werden. Anschließend werden die vermeintlich primitiven und kurzen Textnachrichten auf der für den IS wichtigsten Kommunikationsplattform Twitter ins Blickfeld genommen. Die genauere Betrachtung von Twitter ist dabei besonders interessant, da dort neben Gruppenstrukturen, die im daran anschließenden Teil betrachtet

werden, auch Propagandastrategien messbar werden. Neben mehreren Regierungen und anderen Institutionen steht diese Propaganda auch im Fokus der Plattformbetreiber, wie das Beispiel Twitter zeigt. So wurden bis August 2016 bereits über 360 000 IS-Profile gesperrt, wenngleich die Verbreitung der Propaganda dadurch jedoch nicht gestoppt werden konnte (Klausen, Marks, & Zaman, 2016, S. 2).

Konkret in Bezug auf Twitter bietet sich des Weiteren der Vorteil, dass diese Plattform von verschiedensten Personen mit unterschiedlichem Alter, Geschlecht und Bildungsstand genutzt wird und die dort ihre Meinung verbreiten, weshalb Twitter durchaus als wertvolle Quelle genutzt werden kann. Außerdem ist gerade das Vorhandensein von großen Textmengen ein großer Vorteil, der sich zum Beispiel für korpuslinguistische Verfahren nutzen lassen kann und weitreichende Erkenntnisse liefern kann. Dadurch, dass Twitternutzer aus verschiedensten Ländern kommen, eröffnet sich sogar die Möglichkeit, Datensätze aus spezifischen Ländern, beziehungsweise in verschiedenen Sprachen zu generieren, was weitere spezielle Studiendesigns ermöglicht. Gerade für die Analyse von gruppenspezifischen Inhalten ergibt sich dadurch eine höchsteffektive Möglichkeit der Analyse von Interessen oder Ideologien, die sich gar nach geographischen Orten unterscheiden lassen (Mansour, 2018, S. 96f). Auch ist der Aspekt der zeitlichen Bereitstellung von Informationen ein Punkt, der besondere Aufmerksamkeit verdient. Beispiele wie das Erdbeben in Japan 2011 oder der arabische Frühling haben bereits in der Vergangenheit gezeigt, dass Informationen über soziale Medien wie Twitter in Echtzeit bereitgestellt werden und damit schneller öffentlich zugänglich sind, als es bei traditionellen Medien der Fall ist. Diese in kurzen Nachrichten verpackten unstrukturierten Informationen zu strukturierten Informationen umzuwandeln, beziehungsweise diese überhaupt zu erfassen, ist eine große Herausforderung der Informationsextraktion und damit auch

des OM und in diesem Themenkomplex von ständiger Bedeutung (Piskorski, & Yangarber, 2013, S. 41f).

2.1 Forschungsstand im Kontext der Online-Terrorismusbekämpfung

Wie bereits angedeutet, gibt es eine große Anzahl von Forschungen, die eng mit dem Ansatz dieser Arbeit verbunden sind. Im Folgenden wird der Forschungsstand des Themenkomplexes jedoch eher methodisch als chronologisch vorgestellt, da der Großteil der Arbeiten in den letzten Jahren veröffentlicht wurde und eine methodische Unterteilung daher sinnvoller ist. Zunächst wird dabei auf Werke eingegangen, die sich aus verschiedenen Perspektiven heraus mit dem IS beschäftigt haben. Anschließend werden – methodisch sortiert – Arbeiten beschrieben, die experimentelle Ansätze mit verschiedenen Methoden verfolgen und zu messbaren Ergebnissen kommen, wie es auch diese Arbeit beabsichtigt.

Ein Schlüsselfaktor für CT ist das Verständnis, wie die Struktur der betrachteten Gruppen aufgebaut ist. Einen wichtigen Beitrag hat diesbezüglich Pearson (2018, S. 868f) geleistet, die das Verhalten von Personen, die dem IS positiv zugeneigt sind, im Falle einer Sperrung ihrer Twitterprofile untersucht hat. Dabei fand sie heraus, dass eine Sperrung des Profils keineswegs nur als Verbindungsunterbrechung, sondern vielmehr als gruppenbildendes Ereignis zu sehen ist. Dafür suchte sie je fünf Ausgangsprofile für Männer und Frauen, anhand derer weitere Profile gefunden wurden, untersuchte Hashtags, die geschlechtsspezifisch von IS-Sympathisanten genutzt werden und beobachtete diese Profile im Hinblick auf das Nutzerverhalten (Pearson, 2018, S. 856f). Eine ähnliche Studie haben Huey, Inch und Peladeau (2019, S. 450) durchgeführt, die über den Zeitraum eines Jahres 93 Twitterprofile von Frauen beobachtet und die Gruppenstruktur im Hinblick auf weibliche Aktivitäten in der Internet-Community analysiert

haben. Dabei haben sie acht verschiedene Rollen mit unterschiedlichen Funktionen und Verhaltensweisen herausgearbeitet, die weibliche IS-Sympathisanten einnehmen können. Interessant ist das Ergebnis, dass Frauen in der Gruppe eine eher unterstützende Funktion ausfüllen, wenngleich auch manchmal typisch männliche Rollen wie Rekrutierer oder gar Terrorist eingenommen werden (Huey, Inch, & Peladeau, 2019, S. 458). Auch Awan (2017, S. 142f) hat eine Analyse der sozialen Medien im Hinblick auf Akteure und Gruppenstruktur durchgeführt, die die Online-Strategie des IS beleuchtet. Neben den Absichten, die die Terrorgruppe hat, wurden auch sieben Charaktertypen definiert, nach denen die beobachteten Nutzer auf Twitter und Facebook klassifiziert wurden. Dabei wurde herausgefunden, dass gut 90% der Nutzer männlich sind und aus völlig unterschiedlichen Ländern stammen, zu denen sowohl westliche Länder (UK, USA, Australien und Deutschland), als auch muslimische Länder (Pakistan, Indonesien, Ägypten, Saudi Arabien, Türkei und Libyen) gehören (Awan, 2017, S. 146).

Ein weiterer wichtiger Faktor ist die Fähigkeit des IS, die eigens produzierte Propaganda wirkungsreich zu verbreiten. Auf diesen Aspekt hat sich die Arbeit von Melki und Jabado (2016, S. 94) fokussiert, die die jihadistische Propaganda in Form von Videos, Magazinen und Online-Propaganda aus mehreren Perspektiven heraus analysiert und bewertet hat. Dabei fanden sie heraus, dass der höchstprofessionelle Einsatz von Terror, sozialen Netzwerken, effektiven Storytelling-Techniken und Branding essentiellen Zielen dient, zu denen vorrangig die Erweiterung der medialen Reichweite und die Errichtung eines Gruppenbildes gehören, um weltweites Publikum zu erreichen und neue Unterstützer zu generieren. Auch die Arbeit von Richards (2016a, S. 208f) hat die Wirkung der Propaganda des „Islamischen Staates“ betrachtet, allerdings aus einer politisch-ökonomischen Perspektive heraus. Er stellte dabei fest, dass al-Qaidas Ansatz der ideologischen „sales“ und „franchises“ ähnlich wie westliche Marketingstrategien organisiert werden,

beziehungsweise der IS großen Wert auf ideologisches Branding legt, was letztlich beabsichtigt, politisches, kulturelles und ökonomisches Kapital zu akkumulieren. Zu diesen Gemeinsamkeiten gehört auch Merchandising, was häufig in Indonesien oder der Türkei produziert und im Internet verkauft wird (Richards, 2016a, S. 210). Auch Novenario (2016, S. 957) hat sich mit der Propagandawirkung des IS beschäftigt, indem sie veröffentlichte Magazine von AQ und IS inhaltsanalytisch auf fünf verschiedene Faktoren hin untersuchte, zu denen Attrition, Intimidation, Provocation, Spoiling und Outbidding gehören. Sie fand heraus, dass von AQ veröffentlichte Magazine vor allem auf Propaganda beruhen, die Attrition (Zermürbung), aber auch zu etwas geringeren Teilen Outbidding (Überbietung) und Spoiling (Verderben, Beeinträchtigung) vermitteln, während Provocation (Provokation) und Intimidation (Einschüchterung) bei der Propaganda von AQ eine geringere Rolle spielen. Bei der Propaganda des IS liegt der Fokus vor allem auf der Vermittlung von Outbidding, aber auch zu leicht geringeren Anteilen von Intimidation, Spoiling und Attrition. Provocation hingegen spielt bei der Propaganda des IS ebenso wie bei der von AQ eine geringere Rolle. Interessant ist auch die jeweilige Gewichtung der einzelnen Magazine, die die Veränderungen der Propagandastrategien über die Zeit verdeutlichen (Novenario, 2016, S. 958). Der Ansatz, eine Resonanzanalyse zur Messung der Popularität von IS-Propaganda in einem lokalen Gebiet durchzuführen, wurde von Marcellino, Cragin, Mendelsohn, Cady, Magnuson und Reedy (2017, S. 32f) verfolgt, wobei bewusst Ägypten als Versuchsregion ausgewählt wurde, da dort sowohl säkulare, als auch moderate und extremistische islamische Gruppen um die Deutungshoheit im Land kämpfen. Die Resonanzanalyse ist dabei ein Instrument auf Basis korpuslinguistischer Methoden, um die Effektivität der Verbreitung von Propaganda und die Aufnahme dieser durch die Bevölkerung messbar zu skalieren und auf geographische Einheiten zu reduzieren. Auf Grundlage dieser Analyse wurde

gezeigt, dass nur ein geringer Teil der Nutzer in den vier betrachteten Regionen in Ägypten die IS-Propaganda positiv aufnimmt, wenngleich die Unterstützung massiv ausfällt, sofern sie denn vorhanden ist. Des Weiteren fiel auf, dass der IS seine Popularität in der Region Sinai seit 2014 steigerte (Marcellino et al., 2017, S. 51).

Die Online-Strategie des IS ist dabei stets das wichtigste Mittel und steht demzufolge auch im Hauptfokus des CT. Um in sozialen Medien entsprechende Akteure zu finden, haben Cohen, Johansson, Kaati und Mork (2014, S. 254) auf die vielseitigen Möglichkeiten der Textanalyse aufmerksam gemacht. Dabei verfolgten sie vor allem den Ansatz, linguistische Marker in Texten zu finden, die darauf hinweisen könnten, dass es sich um einen potentiellen Nutzer der Zielgruppe handeln könnte. Ihnen zufolge sind dabei vor allem der Gebrauch von bestimmten Adjektiven in Bezug auf Gruppen oder Personen von Bedeutung, da solche Zielpersonen (ZP) häufig intensiver auf positive und negative Ereignisse, die ihre Gruppe betreffen, reagieren. Beispielhaft wurde in diesem Zusammenhang die Analyse von AQ-Transkriptionen genannt, die mithilfe korpuslinguistischer Software untersucht wurden und verdeutlichten, dass bestimmte Wörter, aber auch Satzkonstruktionen und Ähnlichkeiten im Sprachgebrauch messbar sind und als Vorlage zur Identifikation ähnlicher Texte oder Schreibstile dienen können. In einer anderen Arbeit von Cohen et al. (2018, S. 154) wurde vergleichbar verfahren und auf dem Prinzip der linguistischen Marker aufgebaut. Anhand einer Analyse von Redetranskriptionen von AQ in Afghanistan (AQA) und AQI wurden ideologische Unterschiede zwischen den beiden Gruppen anhand ihres Sprachgebrauchs extrahiert, indem mithilfe einer Kategorisierung linguistische Marker aus diesen Transkriptionen identifiziert und bewertet wurden. Mithilfe dieses Verfahrens zeigten Cohen et al. (2018, S. 156f) nicht nur ideologische Unterschiede zwischen den Akteuren (in diesem Fall Bin Laden und Al-Zawahiri), sondern auch

zwischen den regionalen Gruppen (AQA und AQI) und haben darauf aufmerksam gemacht, dass dieses Verfahren als ein Mittel zur Terrorismusbekämpfung dienen kann. Einen ähnlichen Ansatz verfolgten auch Richey und Binz (2015, S. 354f), die neben weiteren Methoden vor allem für eine Inhaltsanalyse von Texten plädierten, die drei Analysedimensionen beinhaltet. Dazu gehören eine Wort-, eine Themen- und eine Objektanalyse, um identifizier- und suchbare Indikatoren in Texten zu finden. Außerdem verdeutlichten sie die Vorteile einer Authorship Analysis, die anhand unterschiedlicher Faktoren zu hervorragenden Ergebnissen bei der Ermittlung von Identitäten führt und besonders vielversprechend beim CT eingesetzt werden kann. Anhand einer Authorship Analysis lässt sich der Schreibstil eines Individuums anhand mehrerer Indikatoren analysieren. Besonders gut geeignet sind lexikalische, syntaktische, strukturelle und inhaltspezifische Kriterien sowie andere Eigenheiten, die dabei helfen, den Stil eines bestimmten Individuums in einem bestimmten Textstück nachzuweisen (Richey, & Binz, 2015, S. 354f). Auch Brynielsson, Horndahl, Johansson, Kaati, Mårtenson und Svenson (2013, S. 3) verfolgten einen verwandten Ansatz, indem sie verhaltensbasierte Marker in Texten definierten, anhand derer sich LW identifizieren und auch Alias Matching realisieren ließen. Zu diesen verhaltensbasierten Markern gehören nach ihrem Analysemodell Leakage, Fixation und Identification, die als Indikatoren dienen können, um eine mögliche Terrorabsicht eines LW zu erfassen. Dabei wird sich der Umstand zunutze gemacht, dass fast alle LW Teil von Online-Foren oder anderen Formen von sozialen Netzwerken sind (Brynielsson et al., 2013, S. 1). Wie auch bei den vorher genannten Arbeiten fokussieren sich die Analysen hierbei auf den Gebrauch von Sprache, indem beispielsweise in der Leakage-Dimension auf Formulierungen fokussiert wird, die eine Absicht beinhalten. Analog dazu wird in der Identification-Dimension der Fokus auf positive, beziehungsweise negative Formulierungen in Bezug auf eine Gruppe

gelegt, während bei Fixation die Häufigkeit der Nennung eines Subjektes betrachtet wird. Zurückgegriffen wurde dabei stets auf Instrumente der Korpuslinguistik, um sowohl statistische Daten zu generieren, als auch semantische Relationen herstellen zu können (Brynielsson et al., 2013, S. 8f). Neben dem textanalytischen gibt es aber auch einen nutzerzentrierten Fokus. Die Arbeit von Klausen, Marks und Zaman (2016, S. 3f) hatte zum Ziel, Extremisten im Internet anhand ihres Online-Verhaltens zu identifizieren. Dabei formulierten sie ein Modell, dass auf verschiedenen Kriterien beruht, zu denen Sperrungen, die Erstellung von mehreren Profilen, das erneute Finden und Anfreunden nach einer Sperrung und das Finden von gesperrten Nutzern gehören. Dafür beobachteten sie fast 650 000 Profile, von denen in der Zeit von Juni bis September 2015 über 35 000 Profile zeitweise gesperrt waren und auf mehrere Fragestellungen hin untersucht wurden (Klausen, Marks, & Zaman, 2016, S. 6). Besonders interessant sind die Ergebnisse, die das Nutzerverhalten nach einer Sperrung des Profils skizzieren. So observierten die Forscher Profile von einigen Nutzern, die weit über 20-mal den Profilnamen geändert haben, aber dennoch einer wiederholten Sperrung nicht entgehen konnten (Klausen, Marks, & Zaman, 2016, S. 10f). Weitere bedeutende Ergebnisse waren die wirksame Demonstration von Methoden, die Profilähnlichkeiten für das Alias Matching berechnen sowie ein Refollowing Model, das weitreichende Möglichkeiten für CT eröffnet (Klausen, Marks, & Zaman, 2016, S. 18f). Benigni, Joseph und Carley (2017, S. 13f) haben auch eine Reihe ähnlicher Analysen durchgeführt, bei denen Dynamiken im Rahmen von sozialen Netzwerken betrachtet wurden. Dabei wurden nicht nur die verwendeten Sprachen betrachtet, sondern auch thematische Klassifizierungen vorgenommen. So konnte gezeigt werden, dass ein Teil der sehr konservativ eingestellten arabischsprachigen Nutzer häufig Kanälen folgt, die muslimische Theologie thematisieren, wenngleich der Jihad nicht offensiv unterstützt wird. Auch konnte bewiesen werden, dass

einige Propagandisten bewusst Computer-Bots benutzen, um den eigenen Einfluss zu steigern und die Popularität zu verstärken. Neben diesen Erkenntnissen wurde auch ein genauerer Blick auf viele populäre, von IS-Sympathisanten genutzte Hashtags geworfen, womit Benigni, Joseph und Carley (2017, S. 17) zeigten, dass das gegenseitige kopieren und nennen von Profilen (retweeten) ein wichtiger Teil der IS-Online-Strategie darstellt. Beispielsweise wurden Hashtags, die über 75 000-mal erschienen, nur effektiv von 100 verschiedenen Nutzern verwendet.

Im Kontext des IS häufig durchgeführte experimentelle Forschungsarbeiten sind auch solche, die auf einer SA beruhen und anhand von Wortgewichtungen die positive oder negative Haltung in einem Text analysieren. Eine in einem politischen Kontext liegende Arbeit verfassten Siegel, Deuschle, Lenze, Petrovic und Starker (2017, S. 68f), die sich mit der Frage beschäftigten, ob Meinungsumfragen von Sozialforschungsinstituten überhaupt noch zeitgemäß sind, wenn es doch mit Twitter ein Medium gibt, auf dem politisch aktiv diskutiert wird. Auf Grundlage dieser Frage wurde mithilfe einer SA die Haltung von Nutzern des sozialen Netzwerkes Twitter zu ausgewählten Politikern über einen Zeitraum von vier Wochen betrachtet und diese Sentimentscore-Kurve später mit dem ZDF-Politbarometer verglichen. Die Ergebnisse zeigen auf der einen Seite, dass eine SA ein probates Mittel zum Monitoring von sich veränderten Meinungsbildern in sozialen Netzwerken darstellen kann, was sich auch durch die großen Ähnlichkeiten in Bezug auf das ZDF-Politbarometer ausdrückt. Auf der anderen Seite zeigten sich auch die Grenzen der automatischen Klassifizierung von Text (Siegel et al., 2017, S. 72f). Einen interessanten Ansatz im Kontext des Terrorismus verfolgten Ghajar-Khosravi, Kwantes, Derbentseva und Huey (2016, S. 83), die eine wörterbuchbasierte Konzeptgleichung zur Erstellung eines Sentimentscores erstellt haben. Angewendet wurde diese Gleichung auf sogenannte „ISIS-Fangirls“ und auf

normale Teenager und zeigte schlussendlich, dass große Unterschiede zwischen den beiden Gruppen in den Sentimentscores bestehen, sodass auch von absolut unterschiedlichen Haltungen der TG ausgegangen werden kann (Ghajar-Khosravi et. al., 2016, S. 84). Auch Mansour (2018, S. 98f) führte eine terrorismusbezogene SA durch, in der Tweets über den IS aus acht verschiedenen Ländern, die aus vier westlichen und vier muslimischen Ländern bestanden, extrahiert wurden, anschließend nach Worthäufigkeiten sortiert und mit Sentimentscores gewichtet wurden. Dabei wurde herausgefunden, dass es in der Beschreibung des IS zwischen den Ländern fast keine Unterschiede gibt. Das zeigte sich sowohl in der Wortwahl, mit der die Terrorgruppe beschrieben wurde, als auch beim Verhältnis zwischen positiven und negativen Wörtern (Mansour, 2018, S. 102). Eine besonders detailreiche Studie haben Macnair und Frank (2018b, S. 447f) durchgeführt, in der gleich alle vom IS herausgegebenen Medienarten mithilfe einer SA untersucht und verglichen wurden. Dazu gehörten sowohl veröffentlichte Magazine, als auch Transkriptionen von Videos sowie Tweets in sozialen Netzwerken. Besonders interessant ist dabei, dass die ersten Magazine des IS, die *IS News* und *IS Report* heißen, einen deutlich niedrigeren Sentimentscore haben, als die nachfolgenden Hochglanzmagazine *Dabiq* und *Rumiyah*, deren Sentimentscores nahe bei null liegen und damit nicht sonderlich extrem in Bezug auf die Wortwahl verfasst sind. Auch der Twitter-Sentimentscore pendelt sich etwa auf den gleichen Werten ein, was jedoch bei den Transkriptionen der Videos nicht zutreffend ist. Dort ist der Sentimentscore dermaßen negativ, dass von einem großen Unterschied zwischen den Propagandamedien gesprochen werden kann. In einer anderen Arbeit von Macnair und Frank (2018a, S. 112f) wurde der Schwerpunkt auf die Magazine *Dabiq* und *Rumiyah* gelegt, auf die eine SA angewendet wurde. Allerdings wurde diese nicht auf die gesamten Magazine, sondern nur auf eine Selektion von relevanten (in diesem Fall die meistfrequentierten) Nomen

beschränkt, um irrelevanten Ballast zu vermeiden. Die Ergebnisse dieses Ansatzes zeigen, dass eine große Konsistenz zwischen den beiden Magazinen besteht (Macnair, & Frank, 2018a, S. 114f). Das zeigt sich insbesondere in der Betrachtung der zehn häufigsten positiven und negativen Terme, die sich bei den beiden Magazinen zu großen Teilen gleichen. Ebenfalls interessant ist das Ergebnis, dass der Term „Allah“, der eigentlich positiv vom IS verwendet ist, einen der negativsten Gewichtungswerte hat. Eine Begründung könnte darin liegen, dass die umliegenden Terme eine äußerst negative Gewichtung veranlassen, wofür in diesem Zusammenhang als Beispiel der diese Arbeit einleitende Satz dienen möge (vgl. S. 6). Eine Betrachtung der Entwicklung der Sentimentscores über die Zeit zeigt zudem, dass sich die Werte der beiden Terme „Islamic“ und „State“ signifikant mit der Zeit verbessert haben. So tauchen beide zwar als zweit-, beziehungsweise drittnegativster Term in *Dabiq* auf, sind allerdings in *Rumiyah* nicht mal mehr unter den ersten zehn negativen Termen (Macnair, & Frank, 2018a, S. 116).

Es zeigt sich also, dass dies ein sehr aktueller Themenbereich ist, dem sich eine Vielzahl von Forschern bereits gewidmet hat, und die aus unterschiedlichsten Perspektiven heraus und mit verschiedenen Methoden diesen Themenkomplex untersucht haben. Durch die Dynamik der Terrorgruppe in Bezug auf Propaganda, Kommunikationsmedien, Gruppenverhalten, Sprachen, unterschiedliche Charaktere und Sympathien durch LW in verschiedensten Ländern ergibt sich eine komplexe Problemsituation, die mit mehrdimensionalen Methoden angegangen werden sollte, um Ergebnisse auf verschiedenen Ebenen erzielen zu können. Die hier betrachteten Forschungen sind vor allem solche, die diese Problemsituation dabei nicht nur aus wissenschaftlicher Perspektive betrachtet, sondern auch aktiv Ergebnisse hervorgebracht haben, die in einem sicherheitspolitischen Kontext von Bedeutung sind, da es sich bei der Online-Terrorismusbekämpfung um ein Thema handelt, das höchstaktuell und für

jeden Staat von höchster Bedeutung ist. Vor allem durch die Auswahl unterschiedlicher Methoden und Kombinationen von Analysen lassen sich somit Ergebnisse erzielen, die neue Erkenntnisse bringen. Des Weiteren spielt in dieser Hinsicht auch die Datengrundlage eine wichtige Rolle, da sich diese in jeder Studie ändert, was der Extraktion von Kommunikationsdaten von sozialen Netzwerken auf Grundlage unterschiedlicher Faktoren geschuldet ist. Dies führt bei jeder durchgeführten Studie zu anderen Datensammlungen, was auch Auswirkungen auf die Ergebnisse hat. Da in dieser Arbeit sowohl bereits behandelte Datensätze wie die Propagandamagazine des IS, als auch völlig neue Datenkollektionen von Nutzern aus sozialen Netzwerken verarbeitet werden, ergibt sich in der Summe ein neues Analysemodell, dass in genau dieser Form – nicht zuletzt aufgrund der Kombination unterschiedlicher Methoden – bisher noch nicht aufgetreten ist und eine Antwort auf die oben genannten Fragestellungen bieten soll, indem messbare Ergebnisse vorgestellt werden.

2.2 Problemstellungen des Monitorings im Internet

Nach Betrachtung der Frage, welche ausgewählten Forschungsarbeiten sich dieser Thematik bereits angenommen und Gebrauch von bestimmten Methoden gemacht haben, wird sich nun dem Kontext zugewendet, in den sich diese Arbeit eingliedert. Da bei der Nutzung solcher Methoden – nicht zuletzt aufgrund der Komplexität des Themengebietes – eine ganze Reihe unterschiedlicher Schwierigkeiten auftreten, wird sich im Folgenden verschiedenen Problemstellungen gewidmet, die im Rahmen von CT auftreten und die bei der Nutzung technisch-digitaler Monitoring-Methoden Berücksichtigung finden sollten. Dabei werden vor allem zwei verschiedene Dimensionen berücksichtigt, auf deren Basis diese Herausforderungen beruhen. Einerseits existiert eine technische Dimension von Problemstellungen, die betrachtet werden muss und die eine Art

Einstiegshürde in diesen Themenkomplex darstellt. Auf der anderen Seite gibt es eine linguistische Problemdimension, die die Datenanalyse erschwert und aufgrund der Sprachvielfalt und Sprachkomplexität noch schwieriger zu lösen ist, als es bei Problemstellungen der technischen Dimension der Fall ist. Diese beiden Hauptdimensionen werden im Folgenden betrachtet, um ein Grundverständnis dafür zu erschaffen, mit welcher Schwierigkeit das effiziente OM von Terroristen und deren Sympathisanten behaftet ist. Dabei werden auch Beispiele genannt, um Problemstellungen greifbar und für zukünftige Arbeiten bewusst zu machen. Des Weiteren wird dadurch, dass die Problemdimensionen in diesem Teil ausführlich betrachtet werden, auch das Verständnis für Schwierigkeiten erleichtert, die im späteren praktischen Teil dieser Arbeit aufgetreten sind. Besonders wichtig in diesem Kontext ist die Tatsache, dass nicht jede dieser Problemstellungen unabhängig und für sich existiert, folglich also einzeln gelöst werden kann, sondern dass sich durch die Auswahl verschiedener Methoden jeweils neue Kombinationen von Schwierigkeiten ergeben, die in jedem Kontext und bezogen auf jede Fragestellung individuell gelöst werden müssen – analog zu der vielfältigen Auswahl der Datengrundlage. Eine Betrachtung der technischen und linguistischen Problemdimensionen ist damit eine wichtige Basisvoraussetzung, die bei der Durchführung solcher Analysen anfällt.

2.2.1 Technische Dimension

Das Vorhandensein von technischen Problemen ist kein Alleinstellungsmerkmal des OM, sondern betrifft fast jeden informationstechnischen Bereich. Da jedoch im Kontext von CT einige Besonderheiten existieren, lohnt sich ein Blick auf ebendiese. Im Gegensatz zur gewöhnlichen Nutzung von informationstechnischen Geräten, bei denen die technische Problemdimension eher aus Software- oder Nutzerproblemen besteht, geht es beim OM vor allem um den Zugriff auf relevante externe

Datenquellen, die User-generated-content (UGC) einer ZP beinhalten, worin bereits eine erste (zumindest moralische sowie juristische) Schwierigkeit am Einstiegspunkt dieses Themenkomplexes existiert.

Dieser UGC zeichnet sich zudem durch extreme Flüchtigkeit aus, was starken Einfluss auf die Datenquellen, deren Verfügbarkeit und – besonders im Kontext von Forschungen – die Nachweisbarkeit der Daten hat. Des Weiteren ist dieser Inhalt hochgradig personalisiert und wird häufig nur durch spezifische Zugriffe von einzelnen Nutzern überhaupt erst zugänglich gemacht, was wiederum zur Folge hat, dass viele dieser Inhalte für die Öffentlichkeit nicht nutzbar oder sichtbar sind (Naab, & Sehl, 2014, S. 129). Außerdem gestaltet es sich als so gut wie unmöglich, Zugriff auf alle Daten zu erhalten, die sich beispielsweise im Deep Web befinden, da diese von gängigen Suchmaschinen nicht indexiert sind, zumal die Datenmengen häufig so riesig sind, dass eine manuelle Analyse dieser Daten aussichtslos ist (Brynielsson et al., 2013, S. 2). Konkret am Beispiel des IS zeigt sich sehr gut, dass UGC in sozialen Netzwerken im Gegensatz zu anderen Medien sehr flüchtig ist und, wie beispielsweise die häufig im Deep Web veröffentlichten Propagandamagazine zeigen, durch hohe Zugriffshürden eine gewisse Unsichtbarkeit ausnutzt. Außerdem zeichnen sie sich durch aperiodisches Auftreten und Verschwinden aus, was eine weitere Analyse ebenfalls erschwert. Es zeigt sich also, dass allein aufgrund der Beschaffenheit der Analyseumgebung und der Zieldaten sowie deren Beschaffung bereits die Implementierung einer Strategie für CT eine Schwierigkeit darstellt.

Eine andere Form von Problemen beim OM von potentiellen ZP besteht in deren Verhalten. Da sich Nutzer, die sich als IS-Protagonisten oder Sympathisanten in sozialen Netzwerken betätigen, der Tatsache bewusst sind, dass dies eine für Sicherheitsbehörden relevante Handlung ist, werden von ihnen bewusst Hürden aufgebaut, um eine Verfolgung oder Identifizierung des eigenen Profils zu verhindern oder zumindest zu erschweren. Diese

Hürden können beispielsweise dadurch erreicht werden, dass ein IS-Akteur mehrere Profile benutzt, um selbst im Falle einer Sperrung eines Profils durch die Nutzung von anderen Profilen weiterhin Zugriff auf das soziale Medium zu haben. Ein Alias Matching ist in solchen Fällen ein wirksames Mittel, um auf Grundlage von Ähnlichkeiten im Profilnamen, in Netzwerken oder des Schreibstils mehrere Profile einer Person zuordnen zu können (Brynielsson et al., 2013, S. 10f). Das Beispiel Twitter zeigt, wie Plattformbetreiber ihren Beitrag zum Sperren von Profilen leisten, die extremistische Gruppen unterstützen. Twitter richtet extra Teams von Mitarbeitern ein, die Profile auf terroristische Inhalte untersuchen, die von anderen Nutzern gemeldet wurden. Wenngleich diese Strategie die Reichweite von Terrorgruppen wie den IS beschränkt, so stellt sie dennoch keine Lösung für dieses Problem dar, da die Nutzer solcher Profile immer wieder nach einer Sperrung neue Profile erstellen, alten Freunden folgen und erneut propagandistische Inhalte verbreiten. Ein weiteres Problem besteht darin, dass der Nutzen von Profilsperren endet, sobald die Nutzer auf andere Plattformen oder soziale Netzwerke ausweichen, auf denen sie nicht von einer Sperrung betroffen sind (Klausen, Marks, & Zaman, 2016, S. 2f). Wie Pearson (2018, S. 865) bereits zeigte, stellt eine Sperrung des Profils ein bedeutendes Ereignis im Lebenszyklus eines Profils von IS-Sympathisanten auf Twitter dar, das fast rituell und von gut 75% aller dieser Nutzer zelebriert wird. Für gewöhnlich folgt auf eine Sperre die Rückkehr mit einem neuen Profil, das wiederum anschließend von anderen Profilen unterstützt wird, um sich neu aufzubauen. Auch die Existenz von Back-up-Profilen stellt in diesem Zusammenhang das OM von IS-Sympathisanten vor große Schwierigkeiten (Pearson, 2018, S. 867). Ein weiteres essentielles Problem ist, dass das Online-Verhalten von IS-Sympathisanten nicht in jedem Fall und immer einen strafbaren Verstoß gegen geltende Gesetze oder die Twitter-Richtlinien darstellt. Auf diese Weise stellt zum Beispiel das bloße Teilen oder Antworten auf

propagandistische Inhalte zunächst erst einmal keinen Rechtsbruch dar, wenngleich es im Sinne der IS-Propagandastrategie eine bedeutende Rolle spielt, da es eine Verbreitung der Terrorpropaganda darstellt. Wenn im darauffolgenden Schritt die Konversation zu einer anderen Kommunikationsplattform verlagert wird, ist der positive Effekt der Sperrung von IS-Profilen vollständig verloren (Benigni, Joseph, & Carley, 2017, S. 2). Ein Modell von Striegher (2013, S. 45) skizziert anschaulich diese Problemstellung, die im Rahmen der technischen Problemdimension besteht (Abb. 1). In diesem Modell werden vier qualitative Ebenen von Informationen definiert, die für Sicherheitsbehörden von großer essentieller Bedeutung sind und im Sinne des CT eine gewichtige Rolle spielen.

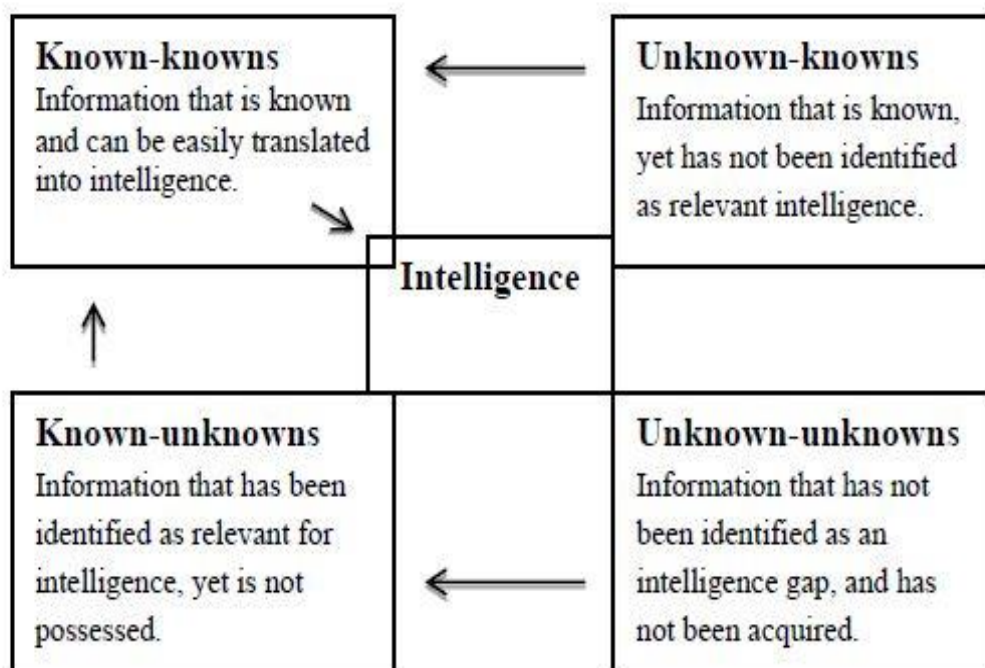


Abbildung 1: Informationsfluss auf dem Weg zur Intelligenz (Striegher, 2013, S.45)

In diesem Modell wird schematisch gezeigt, dass es Informationen gibt, die bekannt sind und auf deren Basis ein Einsatz im Sinne von geheimdienstlicher Intelligenz vorgenommen werden kann (Known-knowns). Dies bildet im

Prinzip den optimalen Zustand ab, da die Informationen bereits vorhanden sind und als relevant gewichtet wurden. Es gibt allerdings auch solche Informationen, die bereits vorhanden sind, allerdings noch nicht als relevant für das jeweilige Informationsbedürfnis identifiziert worden sind (Unknown-knowns). Anhand dieser Informationen ließen sich, sofern diese Ressourcen genutzt würden, weitere wertvolle Möglichkeiten für CT erschließen. Deutlich schwieriger gestaltet sich die untere Ebene des Modells; angefangen bei Informationen, die als relevant bewertet werden, sich aber noch nicht in eigenem Besitz befinden (Known-unknowns). Diese müssen im Regelfall auf irgendeine Weise erschlossen werden, um genutzt und in den optimalen Zustand (Known-knowns) transferiert werden zu können. Die schwierigste Situation ergibt sich, wenn es Informationen gibt, die nicht als relevant bewertet wurden, obwohl sie es sind und sich außerdem nicht in eigenem Besitz befinden (Unknown-unknowns). Dies ist der suboptimale Zustand, den es aus Sicht der Sicherheitsbehörden stets zu verhindern gilt.

Um die technische Problemdimension beim OM anhand dieses Modells zu skizzieren, wird zunächst ein Blick auf die Known-unknowns geworfen. In dieser Ebene befinden sich vor allem solche Informationen, die durch Hürden vor Zugriff geschützt sind, sich im Deep Web befinden, unsichtbar sind oder in einer riesigen Datenmasse verschwimmen. IS-Protagonisten und deren Sympathisanten siedeln sich eher in der Ebene der Unknown-unknowns an, wenngleich viele dieser Personen nach einiger Zeit zu Known-unknowns werden, da sie durch ihr Verhalten in den Fokus der Sicherheitsbehörden geraten. Die Ebene der Unknown-knowns spielt hingegen auch eine wichtige Rolle für CT, da sich auf Grundlage dieser Informationsebene Daten, die zum Beispiel in sozialen Netzwerken zu finden sind und sich im Rahmen von SOCMINT extrahieren lassen, Hinweise über potentielle IS-Protagonisten liefern und diese identifizieren könnten. Abschließend lässt sich in Bezug auf die Informationsebene der Known-knowns außerdem noch anmerken, dass

von hier aus weitere Methoden des OM dazu führen können, ZP zu überführen oder potentielle terroristische Aktivitäten im Vorfeld zu vereiteln. Es zeigt sich also im Rahmen der Betrachtung der technischen Problemdimension, dass eine Vielzahl von Zugangshürden existiert, die zunächst einmal überwunden werden müssen. Wenngleich viele effektive Methoden für das OM existieren, so steht doch zunächst immer die Identifikation von Zielobjekten im Vordergrund. Selbst im Falle einer Identifizierung stellen sich aber aufgrund des Nutzerverhaltens dieser ZP weitere Folgeprobleme ein, die nicht außer Acht gelassen werden können. Durch den Wechsel auf andere Plattformen besteht zudem immer das Risiko, dass das OM fehlschlägt und eine potentielle terroristische Aktivität, die sich im Internet angekündigt hat, nicht vereitelt werden kann. Jedoch gibt es neben der technischen auch noch eine linguistische Problemdimension, welche das OM erschwert – und zwar unabhängig von der Zielgruppe oder dem Kommunikationsmedium.

2.2.2 Linguistische Dimension

Neben der technischen existiert auch eine linguistische Dimension von Problemen, die für OM eine wichtige Rolle spielt, da ohne die Beachtung dieser ein erfolgreiches Monitoring unmöglich ist. Essentiell ist an dieser Stelle die Erkenntnis, dass es sich um eine völlig andere Ebene von Schwierigkeiten handelt, als dies bei der technischen Dimension der Fall ist. Während technische Probleme, die bereits behandelt wurden, häufig eine Einstiegshürde darstellen, entstehen durch linguistische Probleme eher Schwierigkeiten im operativen Ablauf des Monitorings. Durch die Komplexität von Sprachen gibt es eine Vielzahl von potentiellen Fehlerquellen, die bei der automatischen Verarbeitung durch informationstechnische Programme auftreten können und unweigerlich zu Fehlinterpretationen führen. Diese Probleme lassen sich zwar durch eine hohe

Qualität der Analysesoftware teilweise ausgleichen, allerdings gibt es aufgrund der Vielzahl verschiedener Anwendungsfelder spezifische Probleme in Bezug auf den Sprachgebrauch, die bei jedem Analyseobjekt auftreten und individuell betrachtet werden müssen. Auf diese Problemstellungen soll im Folgenden eingegangen werden, um ein Verständnis dafür zu erzeugen, welche linguistischen Schwierigkeiten im Rahmen von Arbeiten, die auf Grundlage von korpuslinguistischen oder vollautomatischen Methoden arbeiten, auftreten können.

Bei der vollautomatischen Inhaltsanalyse existieren bereits seit jeher linguistische Probleme in der Hinsicht, als dass Analysesoftware nicht in der Lage ist, die Komplexität der Sprache adäquat zu erfassen und zu verarbeiten. Das zeigt sich deutlich an der Tatsache, dass eine automatische Inhaltsanalyse immer nur in dem Umfang möglich ist, den eine Software zur Verfügung stellt (Mehl, & Gill, 2010, S. 111). Bei allen Formen von Texten steht ein Analyseprogramm vor der Herausforderung, dass der Schreibstil eines Individuums von vielen Faktoren wie Alter, Geschlecht, Bildungsstand oder soziale Herkunft abhängig ist, um nur einige Indikatoren zu nennen, von denen sich alle theoretisch auch über die Zeit ändern können. Dies stellt nicht nur für die automatische Inhaltsanalyse, sondern auch für die Authorship Analysis große Schwierigkeiten dar (Athira, & Thampi, 2015, S. 186). Andere Programme, die einen worthäufigkeitsbasierten Ansatz verfolgen, stehen hingegen in der Kritik, grammatische Konstruktionen völlig außer Acht zu lassen und sich dem Problem aus einer zu minimalistischen Perspektive heraus anzunehmen. So können diese Programme nicht zwischen einfachen semantischen Beziehungen unterscheiden, verstehen keine Metaphern und verarbeiten Ironie wortwörtlich (Mehl, & Gill, 2010, S. 112). Eine ähnliche Schwierigkeit existiert auch für die Informationsextraktion aus Texten oder die Erfassung von Termen für die Entwicklung von Indexierungswörterbüchern, bei der wiederum interessanterweise seit jeher

die Nutzung von worthäufigkeitsbasierten Ansätzen unproblematisch ist, wenngleich diese erfassten Terme anschließend auf ihre Grundform reduziert werden müssen (Lustig, 1986, S. 13f). All diese Ansätze treffen zudem noch auf die typischen Schwierigkeiten, die beim Natural Language Processing (NLP) ohnehin gewöhnlich auftreten.

Eine besonders schwierige Ausgangslage ergibt sich bei der Verarbeitung von Daten, die aus sozialen Medien akquiriert wurden, was nicht nur dem großen Umfang der Datenmenge, sondern auch der bereits erwähnten Flüchtigkeit, Dynamik und Multimedialität geschuldet ist. Aufgrund des vergleichsweise jungen Alters dieser Medien gibt es bisher kaum adäquate Methoden, um die thematische und sprachliche Komplexität der Kommunikation zu erfassen (Scharkow, 2012, S. 28f). Diese sozialen Medien erlauben Nutzern, große Mengen von Kurznachrichten über alle möglichen Themen zu produzieren, was zusätzlich durch den fast barrierefreien Zugang für Nutzer zu diesen Plattformen noch verstärkt wird. Problematisch ist hingegen die Tatsache, dass eine Verarbeitung dieser Textdaten schwieriger zu bewerkstelligen ist, da häufig grammatikalisch nicht korrekt kommuniziert wird und das Textmaterial für gewöhnlich äußerst kurz (Zum Beispiel bei Facebook 255 Zeichen und bei Twitter 140 Zeichen) ausfällt. Die häufigsten grammatikalischen Hindernisse in sozialen Medien sind dabei Satz- und Rechtschreibfehler, falsche Zeichensetzung und unterschiedliche Schriftarten (Piskorski, & Yangarber, 2013, S. 41). Auch aufgrund der Tatsache, dass Sätze zum Beispiel auf der Plattform Twitter nicht standardisiert sind, kommt es auf linguistischer Ebene zu Problemen bei der Verarbeitung von Texten. Zusätzlich wird eine Analyse dieser Textdaten dadurch verkompliziert, dass Umgangssprache, spezielle Ausdrücke, Ironie und Sarkasmus deutlich häufiger als in anderen Textmedien vorkommen und immer problematisch für eine vollautomatische Analyse sind (Siegel et al., 2017, S. 68). In der Praxis zeigt sich, dass komplexe Phrasen und Mehrwortausdrücke viele Programme

ebenfalls vor Probleme stellen, da diese nicht durch eine Kombination der einzelnen Terme erfasst werden können (Bredack, & Lepsky, 2014, S. 2). Mehrwortgruppen haben zudem in unterschiedlichen Sprachen unterschiedliche Erscheinungsformen, was folglich ebenso Auswirkungen auf die Analyseergebnisse hat. Während im Deutschen Mehrwortausdrücke vor allem in Adjektiv-Substantiv-Kombinationen existieren, gibt es diese im Englischen in nominalen Verknüpfungen, was bezogen auf die Verarbeitung durch Analyseprogramme Schwierigkeiten bereiten kann (Kienitz-Vollmer, & Reichardt, 1986, S. 18).

Bei einer SA wird die Komplexität der Sprache ebenfalls schnell deutlich. Anhand einer anschaulichen Sammlung von fehlerhaften Ergebnissen, die auf einem wörterbuchbasierten Ansatz bei der automatischen Extraktion aus sozialen Netzwerken beruhen, lassen sich solche Schwierigkeiten in der Praxis zeigen. Diese Probleme reichen von der Nichterkennung von komplexen Phrasen über fehlerhafte Interpretationen von Metaphern, Wortneuschöpfungen und Ironie bis hin zu falschen Gewichtungen aufgrund von Rechtschreib- oder Tippfehlern (Siegel et al., 2017, S. 70f). Diese Schwierigkeiten lassen sich auch auf einen Kontext beziehen, der terroristische Gruppen wie den IS einschließt. Da aufgrund der Kürze der Tweets der Informationsgehalt äußerst gering ist, kann eine SA, die nur einzelne Tweets analysiert, nur in begrenzter Weise Erkenntnisse bringen, da diese Analysemethode generell auf längere Texte ausgelegt ist. Des Weiteren gibt es Fragestellungen, die allgemein bei der Analyse von Daten aus sozialen Medien auftreten. So ist es zum Beispiel unwahrscheinlich, dass eine gesamte Gruppe durch die Aktivitäten im Internet repräsentiert wird, da nicht jedes Mitglied zwingend die gleiche Rolle einnimmt und zum Beispiel Textmaterial produziert. Außerdem ist bei solchen Methoden stets ein Informationsverlust zu betrachten, sofern es sich um andere Medien als Text handelt, wie eine

Transkription von Videomaterial verdeutlicht, bei der visuelle Informationen verlorengehen (Macnair, & Frank, 2018b, S. 453f).

Ein besonders deutlicher Konflikt auf linguistischer Ebene besteht bei der Gewichtung von Termen nach positiver oder negativer Bedeutung, wie es numerisch bei SA-Programmen und subjektiv bei jedem Menschen geschieht. Besonders anschaulich wird dieser Konflikt bei Khwaileh und Khuwaileh (2012, S. 23f) beschrieben, wie sich bei der Bewertung von bestimmten Termen durch unterschiedliche Journalisten zweier Blöcke (muslimisch und westlich) zeigt. Beispielsweise ist der Begriff „Martyr“ ein bei islamischen Massenmedien höchst beliebter Begriff, der positiv konnotiert ist, während er in westlichen Massenmedien höchstens in Form von Zitationen auftritt. Interessant ist auch die Betrachtung des Terms „Liberators“, der von westlichen Medien höchstfrequentiert genutzt wird und positiv konnotiert ist, während islamische Medien diesen Term selten und wenn überhaupt, dann eher in politischem Kontext nutzen, wie beispielsweise regierungsnahe Medien in Afghanistan. Am deutlichsten wird der Unterschied bei der Nutzung der negativ konnotierten Terme „Invaders“ oder „Colonization“, die von westlichen Medien überhaupt nicht, von islamischen Medien hingegen sehr häufig genutzt werden (Khwaileh, & Khuwaileh, 2012, S. 25f).

Es zeigt sich also, dass Sprache nicht nur höchst komplex und individuell unterschiedlich, sondern auch durch sprachstatistische Ansätze nur teilweise verarbeitet werden kann. Als besonders schwierig stellt sich die Analyse von Daten aus sozialen Medien heraus, da hier neben häufigeren Fehlerquoten und sprachlichen Mitteln wie Ironie und Metaphern auch eine schnelle Dynamik und Flüchtigkeit vorhanden ist, die eine Auswertung des Datenmaterials erschwert. Besonders in Bezug auf die subjektive Wahrnehmung und Bewertung von Termen stellt sich heraus, dass die linguistische Problemdimension eine hochgradig ausgeprägte Schwierigkeit darstellt, die

das OM und alle Bereiche, in denen Aspekte des NLP genutzt werden, vor Herausforderungen stellt.

2.3 Online-Strategie des „Islamischen Staates“

Der „Islamische Staat“ hat sich seit seiner Selbstausrufung im Jahr 2014 zu einer der weltweit bekanntesten Terrorgruppen entwickelt. Neben seiner höchst brutalen Vorgehensweise ist der IS vor allem deshalb gefürchtet, da er innerhalb kürzester Zeit weltweit zehntausende Kämpfer rekrutieren konnte, von denen ein beachtlicher Teil aus westlichen Ländern stammt. Ein entscheidender Faktor für den schnellen Erfolg des IS sind seine Aktivitäten, die sich im Internet abspielen und auf mehreren strategischen Überlegungen basieren, zu denen eine großangelegte Medienkampagne und hochqualitative Propaganda in Form von Magazinen, dem Einsatz von sozialen Medien und Videos gehören (Macnair, & Frank, 2018b, S. 438). Bei diesen Aktivitäten im Internet hat sich der IS stets als flexibel in der Wiederherstellung seiner Online-Präsenz erwiesen, wenn diese unterbrochen oder entfernt worden ist. Besonders deutlich wird dies am Beispiel Twitter, wo Profile der Terrorgruppe oder seiner Sympathisanten nach kurzer Zeit wieder neu erstellt werden, um seine Zielgruppe zu erreichen und die alte Reichweite wiederherzustellen. Diese Plattform wird auch vorrangig mit dem Ziel genutzt, um weltweit neue Mitglieder zu rekrutieren, indem ideologische Propaganda verbreitet wird (Mansour, 2018, S. 97). Dabei wird nicht selten auf geschütztere Kommunikationsplattformen wie zum Beispiel die Messenger-App Telegram ausgewichen, um die Anonymität der Nutzer zu wahren und einer potentiellen Überwachung zu entgehen, um sensible Informationen zu übermitteln (Hamblet, 2017, S.3). In diesem Zusammenhang ist ebenso interessant, ob und wenn ja wie häufig auf peer-to-peer-Messenger verwiesen wird und ob beispielsweise Links oder andere Inhalte verbreitet werden. Dies ist häufig hilfreich bei der Einschätzung,

welche Rolle ein Nutzer im Rahmen einer Gruppe einnimmt (Benigni, Joseph, & Carley, 2018, S. 238).

Mit der Installation eines solchen Medienapparates hat der IS bewiesen, dass er in der Lage ist, höchst effizient neue Kommunikationstechnologien zu nutzen und sich selbst zu vermarkten, wie es bisher noch keine andere moderne Terrorgruppe konnte. Die verbreitete audiovisuelle Propaganda ist dabei bemerkenswert professionell produziert, in mehreren Sprachen verfasst und regelmäßig veröffentlicht, was dem IS eine globale Sichtbarkeit einbringt, die durch den geschickten Einsatz von sozialen Medien noch verstärkt wird (Macnair, & Frank, 2018b, S. 439). Ein weiterer Faktor für die weltweite Aufmerksamkeit ist seine Fähigkeit, bedrohliche Botschaften zu verbreiten, die bei der adressierten Zielgruppe große Wirkung zeigen. Dies zeigt sich insbesondere an der Veröffentlichung von Rekrutierungsvideos nach erfolgten Terroranschlägen oder Gewaltvideos, die ebenso große Verbreitung erfuhren. Wenngleich schon früher Videos von Terrorgruppen existierten, auf denen Enthauptungen zu sehen waren – der IS produziert dieses Material auf einem anderen Qualitätsstandard, was die Nutzung von Storytelling-Elementen und speziellen Kameratechniken zeigt (Melki, & Jabado, 2016, S. 93). Dazu gehört auch, dass die Videos auf westliches Publikum zugeschnitten sind und so produziert werden, wie es westliche Jugendliche stilistisch gewohnt sind. Besonders zeigt sich das in der Verwendung von modernen Produktionstechniken, Grafiken, unterlegten Erzählstimmen und Hintergrundmusik. Interessanterweise basieren jedoch nicht alle Videos ausschließlich auf Gewaltdarstellungen und Verbrechen, sondern auch auf dem idyllischen Leben im „Islamischen Staat“, dem Zusammenleben und der ethnischen Diversität der IS-Mitglieder (Macnair, & Frank, 2018b, S. 441). Durch die Verbreitung in sozialen Medien entfalten diese Videos eine sehr große Wirkung in der Abschreckung seiner Feinde, der Rekrutierung von neuen Mitgliedern und dienen als Waffe der

psychologischen Kriegsführung zur Vollendung der eigenen Ziele, was nicht selten durch die Verbreitung in westlichen Massenmedien ungewollt unterstützt wird (Melki, & Jabado, 2016, S. 93). Ein Beispiel für ein solches Phänomen ist die Enthauptung von James Foley durch den britischen IS-Terrorist „Jihadi John“, von der in den meisten Nachrichtenagenturen weltweit Bilder verbreitet wurden (Richards, 2016a, S. 213).

Die Ziele der Gruppe bestehen vor allem in der Gewinnung von Wiedererkennungswert und Akzeptanz, um das Weltbild des IS in der Welt zu normalisieren. Insofern kann die Online-Strategie des IS auch als ein Mittel der Soft Power gesehen werden, um die eigene Staatsbildung voranzubringen, die sich in Form von Einführung der Scharia, einer religiösen Polizei, religiösen Schulen und anderen typischen staatlichen Funktionen manifestiert. Dies dient nicht nur der Förderung der Akzeptanz in der lokalen Bevölkerung, sondern gleich der der ganzen weltweiten muslimischen Population, um die Werte des IS zu festigen und zu exportieren. Insofern kann es als ein großer Erfolg der Terrorgruppe erachtet werden, dass eine Alternative zu den vielfach korrupten Regierungen in dieser Region geschaffen wurde (Veilleux-Lepage, 2016, S. 45). Dies zeigte sich von Anfang an in der grundsätzlichen Strategie des IS, mediale Aufmerksamkeit zu erzeugen, das Bild einer kraftvollen Gruppe zu portraitieren und dieses an die Zielgruppe weiterzugeben, Unterstützer in anderen Ländern zu finden und das Leben im Kalifat als nachhaltige Alternative zum westlichen Leben darzustellen (Melki, & Jabado, 2016, S. 94). Aus Sicht des CT kann die Annahme, dass eine reine Militärintervention notwendigerweise zur Zerstörung der Terrorgruppe führt, diesbezüglich als ein großer Irrtum bezeichnet werden (Veilleux-Lepage, 2016, S. 46). Besonders deutlich wird dies durch LW, die autark Anschläge verüben können und sich geographisch unabhängig von der Terrorgruppe bewegen, obwohl eine ideologische Verbindung zu dieser existiert. Um einen erfolgreichen Anschlag außerhalb des Territoriums der Terrorgruppe

durchführen zu können, sind eine Reihe von Fähigkeiten und Ressourcen notwendig, die nicht selten über das Internet in Form von Anleitungen bereitgestellt werden, wie das auch beim IS in Form von Magazinen der Fall ist. Eines der populärsten Beispiele ist in diesem Zusammenhang das Online-Magazin *Inspire*, das von AQAP im Internet veröffentlicht wurde. Diese Anschläge von LW sind dabei stets im Sinne einer Terrorgruppe, da es innerhalb der betroffenen Bevölkerung nicht selten zu Anklagen gegenüber den Sicherheitsbehörden kommt, nachdem eine Attacke stattgefunden hat (Brynielsson et al., 2013, S. 3).

Dieses Spiel mit der Angst wird dabei bewusst und zielgerichtet eingesetzt, um die feindliche Seite zu verunsichern und eine bedrohliche Botschaft an die Bevölkerung zu senden. Interessanterweise lassen sich typische Schlüsseleigenschaften, die eine bedrohliche Nachricht in der Theorie definieren, komplett auf die Botschaften übertragen, die vom IS an seine Feinde verbreitet werden: Mit der Botschaft wird eine Absichtlichkeit ausgedrückt, die überzeugend und ergebnisorientiert wirkt; negative Wertigkeit vermittelt dem Zielobjekt, negativen Konsequenzen ausgesetzt zu sein; Kontrolle über das Mittel zum Auftreten des Schadens suggeriert Überlegenheit; eine Forderung könnte zur Abwendung des Schadens beitragen; eine Eventualverbindlichkeit legt die Verantwortung über die Schwere und Wahrscheinlichkeit des Schadens dem Bedrohten zur Last; Glaubwürdigkeit und Bereitschaft zur Aktivierung des Schadens werden mit der Botschaft impliziert; Bedrohungen richten sich in der Regel gegen zukünftige Handlungen (Spitzberg, & Gawron, 2016, S. 46f). All diese Eigenschaften finden sich in den Botschaften, die vom IS verbreitet werden.

Allgemein lässt sich erkennen, dass die Online-Strategie des „Islamischen Staates“ gut organisiert, hochdynamisch und effektiv durchgeführt wird und dabei mehreren Zielen dient, die die Ausbreitung der eigenen Stärke voranbringen sollen. Dabei werden mehrere unterschiedliche

Kommunikationsmedien und Verbreitungswege genutzt, die eine Verfolgung durch OM erschweren.

2.3.1 Propagandamagazine *Dabiq* und *Rumiyah*

Kernelement der Propagandastrategie des IS sind die Online-Magazine *Dabiq* und *Rumiyah*, die in mehreren Sprachen und regelmäßig erschienen sind. Im Rahmen dieser Hefte verbreitet der IS im Hochglanzformat seine Werte, um seine Ziele zu erreichen und direkten Einfluss auf seine Zielgruppe zu nehmen. Im Folgenden werden diese Magazine genauer betrachtet, besonders im Hinblick auf Propaganda, sprachliche Gestaltung und Inhalt. Diese Magazine eignen sich besonders gut, da sie ein Medium sind, das von jihadistischen Terroristen seit den 1980er Jahren genutzt wird (Novenario, 2016, S. 953). Der IS schreibt selbst über diese Magazine, dass sie sich den Themen Einheit, Wahrheitsfindung, Migration, heiliger Krieg und Gemeinschaft widmen. Neben Texten sind dort reihenweise Fotoreportagen, aktuelle Events und andere Themen rund um den „Islamischen Staat“ zu finden. In diesen portraitiert sich der IS als glorreiches neues Kalifat, das große Siege errungen hat und ein romantisches neues goldenes islamisches Zeitalter eröffnet (El Ghamari, 2017, S. 81).

Innerhalb dieser Magazine lassen sich drei wesentliche Elemente finden, die die Grundstruktur dieses Mediums bilden. Als erstes Element sind Artikel zu nennen, die aus längeren Texten bestehen und die sich häufig über mehrere Seiten erstrecken und meist bestimmte Themen, Ereignisse oder juristische Sachverhalte – manchmal auch eine Kombination aus diesen – betrachten. Dabei werden die Texte stets mit hochauflösenden Fotos ergänzt, um diese visuell aufzuwerten. Neben Artikeln sind auch Statements ein häufig zu findendes Element. Diese sind bezüglich der Länge deutlich kürzer – meist nicht mehr als ein paar Sätze oder Absätze lang. In diesen finden sich meist Auszüge aus Werken prominenter historischer oder aktueller Personen oder

Beschreibungen von Fotoreportagen. Als letztes Element neben Artikeln und Statements sind noch die Anzeigen zu nennen, die sich in ihrer Erscheinung nicht von Anzeigen in anderen Zeitschriften unterscheiden, jedoch kurze Aussagen oder Abschnitte aus islamischen Texten beinhalten, die auch häufig von Bildern begleitet werden (Ingram, 2016, S. 467). Durch diese verschiedenen Elemente wird der Leser passiv der Propaganda des IS ausgesetzt und dadurch suggestiv beeinflusst.

Interessant ist die Tatsache, dass die optische Erscheinung des Magazins eine Hochglanzproduktion darstellt und sich der Informationsversorgung nach westlichem Muster annähert. Neben bearbeiteten Fotos von lachenden Kindern, heroisch inszenierten Kämpfern und scheinbarer Idylle sind in diesen Magazinen hochgradig religiös aufgeladene Inhalte zu finden. Durch die Nutzung von Bildern, die Gewaltakte wie Exekutionen oder tote Körper darstellen, sollen beim Leser emotionale Reaktionen ausgelöst werden, wie dies auch in westlichen Medien oder Hollywood-Filmen beabsichtigt ist (Richards, 2016a, S. 215f).

Diese beiden Magazine waren jedoch nicht die ersten, die vom IS veröffentlicht wurden. Bevor es zur Veröffentlichung von *Dabiq* und *Rumiyah* kam, wurden im Mai 2014 zunächst die *Islamic State News* herausgebracht, denen noch zwei weitere Ausgaben im nächsten Monat folgten, und fast zur gleichen Zeit vier Ausgaben des *Islamic State Report*. Diese Magazine waren im Umfang deutlich kleiner und meist auf ungefähr sechs Seiten begrenzt, auf denen hauptsächlich mit Bildmaterial gearbeitet wurde. Erst mit der anschließenden Veröffentlichung von *Dabiq* im Juli 2014 erreichte der IS dann das Niveau, auf dem von einer Hochglanzproduktion die Rede sein kann. In diesen Magazinen wuchs dann auch der Anteil an Textstellen, in denen zum Beispiel militärische Neuigkeiten, Soldatenbiographien, Frauenthemen, Rekrutierungsrufe und religiöse Behandlungen eine Rolle spielten. Dieses Konzept, ergänzt um monatliche

Herausgaben, setzte sich dann auch mit der Veröffentlichung des Magazins *Rumiyah* im September 2016 fort (Macnair, & Frank, 2018a, S. 110f). Diese Magazine spielen in der Rekrutierungsstrategie des IS eine besondere Rolle und tragen einen beachtlichen Teil zu deren Erfolg bei, was in besonderem Maße auch für Frauen gilt, die mit diesem Medium angesprochen werden sollen. So wird die Rolle der Frauen in diesen Magazinen so dargestellt, als sei durch Gemeinschaft, Freundschaft, Heirat und religiöse Ausübung das Leben erfüllender, als es im Westen der Fall wäre. Durch den speziellen Einsatz von sprachlichen Narrativen versucht der IS, eine globale und kollektive Identität für Muslime zu erschaffen, um diese von einer Anhängerschaft zu überzeugen. Dies wird durch ein starkes „Wir-gegen-die“-Narrativ innerhalb der Magazine unterstützt (Macnair, & Frank, 2018b, S. 440). Dabei wird stets darauf Wert gelegt, die korrekte religiöse Auslegung vorzuleben, was mit sehr emotionaler Sprache unterstrichen wird – frei nach dem Motto, auf einer göttlichen Mission zu sein (Macnair, & Frank, 2018a, S. 111).

Interessant ist die Betrachtung der Magazine im Hinblick auf Propagandawirkungen, die beim Leser ausgelöst werden sollen. Dabei dienen verschiedene Faktoren unterschiedlichen Zielen, die mit dem Einsatz dieser erreicht werden sollen. Einer dieser Propagandafaktoren ist Attrition und soll verdeutlichen, dass der IS in der Lage ist, hohe Kosten für seine Feinde zu verursachen und soll auf diese Weise der Staatsbildung zuträglich sein. Ein Beispiel für den Einsatz von Attrition ist die Veröffentlichung eines angeblich von James Foley verfassten Briefes nach seiner Enthauptung, in der dieser die U.S.-Luftschläge für seinen Tod verantwortlich macht, oder die Darstellung von anderen Terroranschlägen (wie in Paris, Berlin oder der Anschlag in einem russischen Passagierflugzeug) als eine Form von Rache. Ein anderer Faktor ist Intimidation, der sich vor allem durch die Exekutionen von politischen Feinden oder Personen mit anderem Lebensstil zeigt und

propagandistisch in den Magazinen verarbeitet wird. Der Faktor Provocation indes wird verhältnismäßig wenig sprachlich abgebildet und findet vor allem dort Verwendung, wo der IS seine schlimmsten Taten behandelt und diese als Rache darstellt. Ein Beispiel dafür ist das Video des jordanischen Kampfpiloten, der lebendig in einem Käfig verbrannt wird, um die Hinrichtung von zwei Jihadisten in Jordanien zu rächen (Novenario, 2016, S. 959f). Spoiling gehört auch zu den genannten Faktoren und zielt vor allem darauf ab, eine Spaltung der Gesellschaft zu erzeugen. Dies geschieht zum Beispiel durch Artikel, in denen behauptet wird, dass ein Moslem dem Jihad beitreten müsse, um ein guter Moslem zu sein. Als letztes ist der Faktor Outbidding als ein elementares Kernelement der Propaganda des IS zu nennen, um nicht zuletzt das selbsternannte Kalifat zu legitimieren und eine Führungsrolle in der globalen jihadistischen Bewegung zu beanspruchen, indem andere Terrorgruppen diskreditiert werden (Novenario, 2016, S. 961f). Über die Zeitspanne von Juli 2014 bis Januar 2016 betrachtet fällt zudem auf, dass Intimidation anfangs ein wesentlicher Kernfaktor der Propaganda war, der jedoch mit der Zunahme von eroberten Gebieten stetig an Bedeutung verlor und stattdessen vom Outbidding abgelöst wurde. Dies geschah nicht zuletzt mit der Absicht, angesichts des wachsenden militärischen Drucks der Anti-IS-Koalition andere jihadistische Kämpfer zum Überlaufen und andere Gruppen zum Treueschwur für den IS zu überzeugen. Wenngleich es ein gutes Argument ist, dass die Brutalität der Terrorgruppe eine Form von Provocation ist, so spiegelt sich diese im Rahmen der Magazine sprachlich nur in geringem Umfang wider (Novenario, 2016, S. 958). Gleiches gilt auch für die sprachliche Reflexion von Spoiling und Attrition.

Jedoch gibt es noch andere Gründe, weshalb diese Magazine überhaupt einen solchen Einfluss entwickeln konnten. Erstens ist es eindeutig, dass diese Magazine offiziell und original von der Terrorgruppe veröffentlicht wurden, was bei der Vielzahl von Inhalten, die täglich im Internet im Namen des IS

produziert wird, nicht immer erwiesen ist. Des Weiteren ist der Inhalt selbst ausgewählt, was eine gewisse Flexibilität in der eigenen Argumentationsstruktur gewährleistet. Drittens ist das Magazin reichhaltig ausgestattet und auf einem hohen Niveau produziert. Außerdem ist der Inhalt strukturiert und folgt einem bestimmten Narrativ, das häufig über mehrere Magazine hinweg aufgeteilt wird und dem Leser so ermöglicht, auch komplexere Themen zu verstehen. Fünftens und letztens ist der Inhalt dieser Magazine höchst einflussreich, was sich an der Entwicklung der Google-Suchanfragen nach einer Veröffentlichung zeigte (Jacoby, 2019, S. 33f). Interessant ist die Betrachtung der Magazine im Hinblick auf religiöse Verweise zum Koran, die in einem Großteil der vorhandenen Kapitel vorzufinden sind und suggerieren könnten, dass durch die häufigen Verweise der Inhalt des Koran halbwegs in den Magazinen abgebildet sei; dies ist jedoch keineswegs der Fall. Stattdessen fokussiert sich der IS auf einen kleinen Teil des Korans, auf den übermäßig häufig verwiesen wird und der in die eigenen Argumentationsmuster passt. In diesem besonders häufig zitierten Teil des Korans geht es vor allem darum, eine wachsende muslimische Gemeinschaft abseits von Mekka innerhalb der neuen Heimat unter Christen und Juden anzuleiten (Jacoby, 2019, S. 34f). Durch die eigenwillige Nutzung dieser Textstellen soll ein vermeintlich religiös belegtes Narrativ erstellt werden, das dem Leser suggeriert, dass der IS die einzig wahre korrekte Leseart des Korans ausübt. Des Weiteren soll auf diese Weise im selbsternannten Kalifat das Monopol des Regimes auf Gewaltausübung, Rohstoffgewinnung und alleinige Autorität im göttlichen Einsatz legitimiert werden (Jacoby, 2019, S. 49f).

Im Allgemeinen zeigt sich, dass die Magazine der Terrorgruppe gut organisiert und hochqualitativ produziert sind, was auch der gezielte Einsatz von Propagandafaktoren zeigt, der beim Leser emotionale Reaktionen auslösen soll. Durch den geschickten Einsatz von Fotos, sprachlichen Bildern

und willkürlichen Verweisen auf den Koran wird ein hochgradig einflussreiches Medium erschaffen, was seine häufig leicht beeinflussbaren Konsumenten wirksam in den Bann zieht und damit der Erfüllung der eigenen gesteckten Ziele dient, indem neue Anhänger rekrutiert werden.

2.3.2 Twitter als Kommunikationsmedium für Sympathisanten

Neben den Magazinen spielt auch die Nutzung von sozialen Netzwerken – vornehmlich Twitter – eine elementare Rolle in der Propagandastrategie des IS. Twitter hat sich dabei im Laufe der Zeit zum Hauptschauplatz der Medien-Kommunikation der Terrorgruppe entwickelt und ist für den IS zur bevorzugten Plattform geworden. Besonders bemerkenswert ist die Effizienz, mit der sowohl lokale Anhänger, als auch weltweit verstreute Sympathisanten mit Echtzeit-Informationen versorgt werden, Rekrutierungen erfolgen, Propaganda verbreitet und die internationale Marke aufgebaut wird. Obwohl es bereits mehrere Versuche gab, die Aktivitäten dieser Akteure einzudämmen, konnten diese bis heute nicht vollständig unterbunden werden (Macnair, & Frank, 2018b, S. 442). Zwar wurde zwischendurch vonseiten der Terrorgruppe versucht, auch andere Plattformen wie VKontakte oder Facebook zu nutzen, jedoch hat sich Twitter als dauerhaftes Kommunikationsmedium des IS durchgesetzt (Weimann, 2019, S. 31).

Dieser Umstand ist auf einige strukturelle Eigenschaften des sozialen Netzwerkes zurückzuführen, die den Nutzern beispielsweise erlauben, anonym zu bleiben, weil es keine strikten Regulierungen gibt. Ein weiterer Grund, weshalb Twitter zum bevorzugten sozialen Netzwerk solcher Gruppen wurde, ist der Informationsfluss, der mit der Nutzung dieses Dienstes erreicht werden kann. Besonders für Bestrebungen im Rahmen des CT spielt Twitter eine zentrale Rolle, da viele Terroristen, die später Anschläge verübten, vorher radikale Inhalte in ihren Profilen veröffentlichten (Ghajar-Khosravi et al., 2016, S. 79). Die Nutzung von Twitter als Kommunikationsmedium

erlaubt dem IS, mit einem globalen Publikum in Kontakt zu treten, das mit einem persönlichen Kontakt niemals erreicht werden könnte. Dies trifft insbesondere auf charakterlich ungefestigte junge Leute zu, die viel Zeit im Internet und besonders in sozialen Netzwerken verbringen und durch die Verbreitung von Propaganda in Form von Videos und Magazinen beeindruckt werden können. Auf diese Weise werden die vom IS verbreiteten Videos, Kurznachrichten, Magazine und Foren zu Radikalisierungswerkzeugen (Awan, 2017, S. 138f).

Die Tragweite dieser Strategie wird deutlich, wenn ein Blick auf die Anzahl von aktiven Profilen geworfen wird, die die Terrorgruppe unterstützen. So wurde geschätzt, dass nach der Selbstausrufung des Kalifats im Jahr 2014 zwischen 46 000 und 90 000 aktive Profile existierten, die sowohl aus offiziellen Propagandaprofilen, als auch aus tausenden inoffiziellen und von Sympathisanten geführten Profilen bestanden. Der dabei veröffentlichte Inhalt wurde in mehreren Sprachen verfasst, wie es auch bei den Propagandamagazinen der Fall ist (Pearson, 2018, S. 850f). Durch eine solche Vielzahl von aktiven Akteuren im Internet ist der IS in der Lage, propagandistische Inhalte zu verbreiten, die zudem nicht wie in Massenmedien gefiltert werden. Des Weiteren passt die dezentrale Organisationsstruktur der Terrorgruppe und insbesondere der Online-Strategie in besonderer Weise zu sozialen Netzwerken, da zum Beispiel OM und CT dadurch deutlich erschwert werden.

In diesen Kontext betten sich auch andere Aktivitäten ein, wie beispielsweise das dezentrale Crowdsourcing, eigens entwickelte Apps und das sogenannte „Hashtag hacking“. Kern dieser Aktivitäten ist die möglichst effiziente Verbreitung von Inhalten, zum Beispiel per „Twitter storm“ (Melki, & Jabado, 2016, S. 96f). Dieser ist sehr genau durchgeplant, gleicht einer informationstechnisch-militärischen Operation und läuft folgendermaßen ab (Abb. 2):

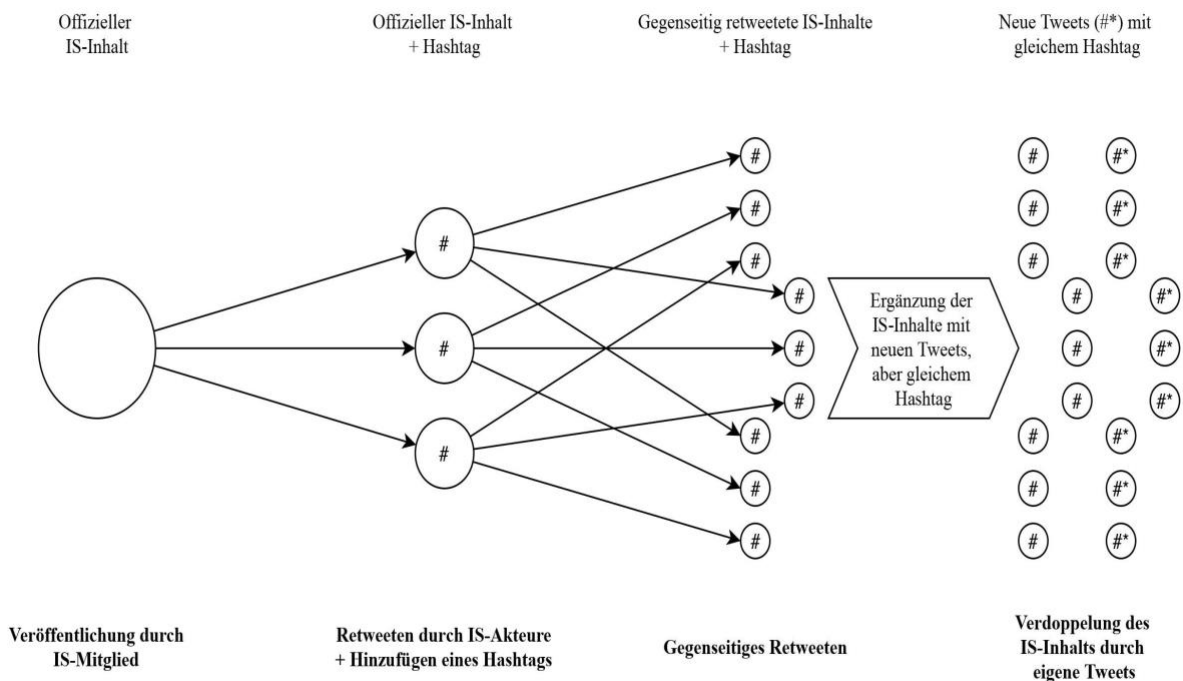


Abbildung 2: *Twitter storm am Beispiel des „Islamischen Staates“* (Eigene Darstellung)

Authentifizierte und von offiziellen IS-Mitgliedern veröffentlichte Inhalte werden von einer weiteren Gruppe von Online-Aktivisten geteilt (retweetet) und mit einem Hashtag versehen. Anschließend werden die gegenseitigen Tweets erneut geteilt und darauf aufbauend neue Tweets geschrieben, die diesen Hashtag beinhalten. Im folgenden Schritt können von anderen Online-Aktivisten diese Inhalte in anderen Plattformen hochgeladen werden, damit die Inhalte im Falle einer Löschung durch eine Plattform noch auf anderen Plattformen auffindbar sind. Dieser Prozess lässt sich sogar auf einer noch größeren Ebene durchführen (Stern, & Berger, 2015, S. 155). Auf diese Weise ist der „Islamische Staat“ in der Lage, den eigens produzierten Inhalt mit einem gewaltigen Faktor zu vervielfältigen und erreicht auf diese Weise mehr Personen aus der Zielgruppe. Neben diesen „Twitter storms“ wurde außerdem eine hauseigene App entwickelt, die eine theologische Verbindung impliziert.

„*The Dawn of Glad Tiding*“ war im Prinzip eine Nachrichten-App, mit der Interessenten über Neuigkeiten im „Islamischen Staat“ informiert werden sollten. Eine Besonderheit dieser App war jedoch, dass es dem IS möglich war, Tweets von den Profilen derjenigen zu senden, die diese App installiert hatten. Dies hatte vor allem zur Folge, dass auf diese Weise die Spam-Filter von Twitter umgangen werden konnten und die eigene Propaganda ungehindert verbreitet werden konnte. Eine weitere populäre Variante war das sogenannte „Hashtag hacking“, bei dem hochgradig populäre Hashtags auf Twitter massenweise von Unterstützern benutzt werden, um ein möglichst großes Publikum zu erreichen. Beispiel dafür ist die starke Nutzung von Hashtags zur Fußball-Weltmeisterschaft #Brazil2014 oder zum Unabhängigkeitsreferendum in Schottland mit den Hashtags #VoteNo oder #VoteYes. Diese wurden dabei stets mit eigenen Hashtags ergänzt, um die eigene Reichweite zu erhöhen (Veilleux-Lepage, 2016, S. 43f). Unabhängig davon lässt sich außerdem feststellen, dass der Freitag der beliebteste Tag der IS-Anhängerschaft ist, um Inhalte zu verbreiten. Diese Taktik ist unter anderem damit zu begründen, dass Tweets über das Wochenende länger haltbar sind und eine Löschung unwahrscheinlicher ist und außerdem der Freitag für Muslime aus religiöser Sicht eine besondere Rolle spielt (Ashcroft et al., 2015, S. 163).

Im Endeffekt fungieren diese Mittel und vornehmlich Twitter als eine Art Megafon, um in Echtzeit Informationen zu verbreiten und die eigene Marke zu unterstützen, wobei einzelne wenige Profile herausstechen. So hatte zum Beispiel @ShamiWitness auf Twitter fast 18 000 Follower bis zu dem Zeitpunkt, als die Person, die mutmaßlich hinter diesem Profil steckte, in Indien verhaftet wurde. Infolgedessen starteten Sympathisanten des IS auf Twitter unter dem Hashtag #FreeShamiWitness eine Protestkampagne gegen dessen Inhaftierung (Awan, 2017, S. 142f). Diese Informationen umfassen indes auch vermeintlich harmlose Aspekte des Lebens im Kalifat, wie am

Beispiel der Hashtags #JihadiFitness oder #LittleMujahid deutlich gemacht werden kann. Durch solche verbreiteten Inhalte soll der Zielgruppe vermittelt werden, dass es sich bei dem selbstverkündeten Kalifat um eine gut funktionierende, harmonische Alternative zur westlichen Lebensweise handelt (Melki, & Jabado, 2016, S. 101). Besonders interessant in diesem Zusammenhang ist jedoch das Phänomen, dass neben den westlichen Produktionstechniken der eigenen Propaganda auch westliche Produkte Teil des Lebens im Kalifat zu sein scheinen. Beispielsweise posierte ein französischer IS-Kämpfer mit einem Glas Nutella. Ein anderes Beispiel zeigt einen Anhänger in einer Adidas-Jacke mit dem Wortspiel „*All Day I Dream About Sharia*“. Am deutlichsten zeigt der Kanal @ISILCats auf Twitter die Absicht des IS, ein möglichst normales Leben im Kalifat darzustellen. Dort werden Kämpfer der Terrorgruppe gezeigt, wie sie sich mit Kätzchen und anderen kleinen Tieren beschäftigen. Untermalt wird dies mit einer verniedlichten Sprache, wie sich am folgenden Beispiel zeigt: Ein IS-Terrorist im Kampfanzug trägt in der einen Hand ein Sturmgewehr und in der anderen eine kleine Katze, was mit dem Satz „*my mewjahid protectz me*“ kommentiert ist (Richards, 2016a, S. 210f). Dieses Beispiel zeigt sehr deutlich, dass es der Terrorgruppe weniger darum geht, der Katze zu unterstellen, ein Gotteskrieger (Mujahid) zu sein, als vielmehr dem brutalen Terror eine harmlose, ja fast niedliche Note zu geben und einen normalen und harmonischen Alltag zu suggerieren.

Mit diesen Inhalten will der IS verschiedene Zielgruppen konfrontieren. Einerseits sollen aktive Gegner und die internationale Öffentlichkeit, andererseits aber auch aktive Mitglieder und potentielle Rekruten oder spezielle Gruppen, wie Missionare oder andere Angeworbene mit diesen Inhalten konfrontiert werden, wobei die Inhalte jedoch je nach spezifischer Zielgruppe unterschiedlich in Bezug auf Wirkung und Zielsetzung ausfallen können (Marcellino et al., 2017, S. 37f). Eines der Primärziele, namentlich die

Rekrutierung, vollzieht sich dabei auf Basis der Radikalisierung, die sich grob in vier Stufen einteilen lässt: Die erste Stufe ist die der „Vorradikalisierung“, in der sich ein potentieller Rekrut befindet, der sich beispielsweise auf der Suche nach eigener Identität befindet. In der nächsten Stufe, der „Selbstidentifikation“, befindet sich dieser Rekrut in einem Stadium, in dem er jihadistische Ideologie als nachvollziehbar und schlüssig erachtet. In der nächsten Stufe, der „Indoktrination“ übernimmt er die Überzeugungen, was nicht selten durch direktes Propagandamaterial verstärkt wird. In der letzten Stufe, der „Jihadisierung“, akzeptiert er sämtliche Mittel des Jihad und wird selbst mehr und mehr dazu bereit, terroristische Aktivitäten auszuführen (Hamblet, 2017, S.6f). Dieser Radikalisierungsprozess ist jedoch nicht eingrenzbar auf bestimmte ethnische, persönliche oder soziale Hintergründe, sondern umfasst sämtliche internationale Gesellschaftsklassen, was nicht zuletzt durch einen hohen Prozentsatz von Konvertiten im IS deutlich wird (Hamblet, 2017, S.5).

Interessant sind einige Verhaltensmuster dieser Akteure in sozialen Netzwerken – vor allem, wenn sie mit einer Suspendierung des Profils konfrontiert sind. Einige dieser Personen, die auf Twitter für die Verbreitung von Propagandamaterial zuständig sind, positionieren in ihren Profilen Merkmale, anhand derer ihre Profile für Follower und Sympathisanten identifizierbar bleiben, was besonders im Falle einer Suspendierung relevant ist. In diesem Fall wird von den gesperrten Akteuren einfach ein neues Profil erstellt, das anhand dieser Merkmale für alte Follower sofort wiedererkennbar sein soll, damit die alte Struktur wiederaufgebaut wird. Besonders macht sich dies bei der Wahl der Profilnamen bemerkbar, die sich nach der Sperrung eines Profils von einem neuen Profil kaum unterscheiden (Klausen, Marks, & Zaman, 2016, S. 9f). Diese Ähnlichkeiten zwischen alten gesperrten und neueröffneten Profilen von IS-Sympathisanten machen sich auch im Rahmen von Screen-Namen, Profilbildern, Freunden und anderen Aspekten bemerkbar

und stellen einerseits die Plattformbetreiber vor Schwierigkeiten, bieten andererseits aber auch für das CT bestimmte Möglichkeiten, da so von diesen IS-Akteuren ungewollt Vergleichsindizes bereitgestellt werden, die ein OM erleichtern (Klausen, Marks, & Zaman, 2016, S. 18f).

Ein Aspekt der Online-Propaganda des IS, der zwar nur eine kleine, aber dennoch ergänzende Rolle spielt, ist die Entwicklung von eigenen Computerspielen, um vornehmlich junge Leute zu rekrutieren. Die Absicht dahinter ist, innerhalb der Computerspieler-Community analog zu den hochprofessionell erstellten Videos die Aufmerksamkeit junger Leute zu gewinnen und Terror als attraktive, fast „coole“ Aktivitäten zu vermitteln. Dafür werden Modifikationen bereits veröffentlichter und populärer Computerspiele – sogenannte Mods – entwickelt und über die bereits angesprochenen Kommunikationskanäle verbreitet. Besonders betroffen von diesen eigens entwickelten Mods sind besonders populäre Computerspiele wie *Call of Duty*, *Grand Theft Auto* oder *Arma III*. Die Hauptänderung dieser Mods ist stets, dass anstatt fiktiver Charaktere westliche Personen, Soldaten des syrischen Regimes oder kurdische Peschmerga getötet werden sollen, wobei die Computerspiele häufig mit jihadistischer Musik unterlegt werden. Auffallend sind auch die Parallelen, die zwischen den offiziell produzierten Videos und den ursprünglichen Computerspielen bestehen, was sich vor allem in der Kameraführung äußert (El Ghamari, 2017, S. 74f). Diese Computerspiele könnten allerdings auch der Kommunikation untereinander dienen, da auf diese Weise keine nachvollziehbaren Spuren entstehen. Eventuell könnten die Pariser Terroranschläge 2015 auf diese Weise geplant worden sein (Richards, 2016b, S. 21). Dies erweitert sich auch auf bei uns bekannte Filme wie *Saw*, *The Matrix*, *American Sniper* oder *V for Vendetta*, die erstaunlich viele Gemeinsamkeiten in der Kamera- und Bildführung mit den offiziellen IS-Propagandavideos aufweisen. Auch hier ist das Ziel wieder deutlich erkennbar – Jugendliche, die in westlichen Ländern leben und oben

genannte Filme oder Computerspiele konsumiert haben, sollen anhand der Motive subtil von diesen bereits bekannten Mustern beeinflusst werden (El Ghamari, 2017, S. 78). Die Nutzung von Twitter ist für den IS also weitaus mehr, als nur ein Kommunikationsmedium mit Sympathisanten; vielmehr ist sie Hauptgrundlage für die schnelle Expansion und Hauptmittel für die Rekrutierung neuer Mitglieder.

2.3.3 Gruppenstruktur und die Rolle der Frauen

Um das höchst organisierte Auftreten der Gruppe in sozialen Netzwerken besser verstehen und zwischen den Akteuren differenzieren zu können, lohnt ein Blick auf die Gruppenstruktur des IS innerhalb des „Online-Kalifats“. Zwischen den verschiedenen Akteuren lässt sich nämlich eine Rollenverteilung erkennen, die im Folgenden genauer betrachtet werden soll. Die Terrorgruppe kann dabei als neo-jihadistische Gruppe definiert werden, was im Prinzip als neue Form des ideologischen Ausdrucks und Militanz verstanden werden kann, die neue Interpretationen der islamischen Theologie mit nicht-islamischen Mitteln der sozialen Organisation und Interaktion kombiniert. Der IS ist dabei in der Hinsicht neo-jihadistisch, als dass er seine eigenwillige Auslegung des Islam und seine Ideologie über soziale Medien verbreitet – also über nicht-islamische Kommunikationsmittel expandiert (Richards, 2016a, S. 206). Kernstück dieser neo-jihadistischen Vorgehensweise des IS, in dessen Auftrag ein jeder Akteur der Gruppe handelt und die die psychologische Grundlage für die Rekrutierungsstrategie darstellt, ist die Dynamik des Radikalisierungsprozesses, die sich „cyclical cognitive reinforcement“ nennt (Abb. 3).

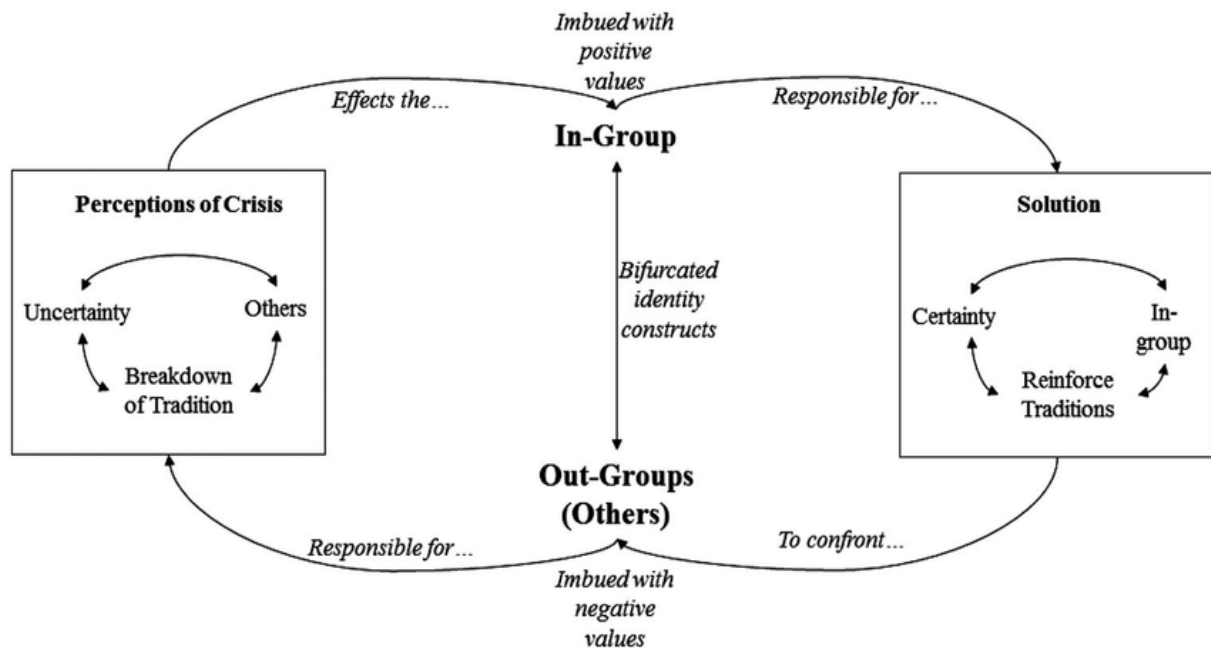


Abbildung 3: *Cyclical cognitive reinforcement* (Ingram, 2016, S.464)

Diese beschreibt im Wesentlichen die Bildung eines bipolaren Gruppenbildes, bei dem die gegnerische Seite für Krisen und Niedergang verantwortlich gemacht wird, dieser Niedergang jedoch nur durch die eigene Gruppe gelöst werden kann. Auf diese Weise soll die gegnerische Gruppe dämonisiert und die eigene gestärkt werden. Die gegnerische Seite (Out-Group) wird dabei mit negativen Eigenschaften beschrieben, die verantwortlich für wahrgenommene Krisen gemacht wird, was wiederum die eigene, positiv dargestellte (In-Group) tangiert. Des Weiteren wird die eigene Gruppe als verantwortlich für die Lösung dieser Problematik dargestellt, um der gegnerischen Seite in konfrontativer Weise zu begegnen. Dieses bipolare Gruppenbild hat somit zum Ziel, die Out-Group mit der Krise zu verknüpfen, während die Lösung dieser Krise nur durch die eigene In-Group realisiert werden kann (Ingram, 2016, S. 464f). Bezogen auf den IS, ist eindeutig in Form der verbreiteten Propaganda und des Gruppenverhaltens in sozialen Netzwerken erkennbar, dass ein bipolares Gruppenbild kreiert werden soll. Dieses Gruppenbild ist jedoch in sozialen Netzwerken nicht immer sofort erkennbar, wie eine kleine

Auswahl von Profilen zeigt, die auf den ersten Blick nichts mit der IS-Ideologie oder einer Gruppendynamik gemeinsam haben. So lassen sich Profile finden, die nur wenige Tweets verfasst haben und sich schwer klassifizieren lassen, während andere Profile der IS-Aktivität und den IS-Mitgliedern folgen, aber selbst kaum Tweets verfasst haben und unter Umständen einfach nur Forschung betreiben. Außerdem gibt es Profile, auf denen alle paar Stunden Koran-Verse veröffentlicht werden – allerdings derart regelmäßig, dass mutmaßlich Twitter-Bots am Werk sind. Wiederum andere Profile verfassen überhaupt keine Tweets, aber folgen Profilen, die IS-Medien verbreiten. Andere Profile veröffentlichen Pornografie in arabischer Sprache, folgen anderen Profilen mit pornografischen Inhalten und haben IS-Mitglieder als Follower (Klausen, Marks, & Zaman, 2016, S. 8). Diese kleine Auswahl von Beispielen zeigt also sehr deutlich, wie groß die Unterschiede zwischen einzelnen Profilen sind, die in unmittelbarer Nähe zu IS-Profilen stehen und unter Umständen selber solche sein könnten.

Viel deutlicher wird die Gruppenstruktur hingegen, wenn die Interaktion untereinander betrachtet wird und auf Grundlage dessen verschiedene Rollen definiert werden, die innerhalb der Gruppe eingenommen werden können. Dabei werden insgesamt sieben verschiedene Gruppentypen deutlich, die sich in ihren Aufgaben teils sehr unterscheiden. Dazu gehören zum Beispiel die „Cyber Mobs“, die soziale Netzwerke nutzen, um eine Mob-Mentalität zu erzeugen und andere dazu drängen, sich für den IS einzusetzen. Dies wird häufig durch Hass-Kommentare, Videos oder andere IS-verwandte Medien durchgeführt. Außerdem gibt es die „Loners“, die sich dadurch auszeichnen, dass sie individuelle Inhalte verfassen und von der IS-Propaganda fasziniert sind, aber eindeutig aus persönlicher Unzufriedenheit handeln. Des Weiteren spielen auch „Fantasists“ eine Rolle, die soziale Medien nutzen, um völlig realitätsfremd über den IS zu philosophieren, aber teilweise direkte Wünsche äußern, doch bitte für die Terrorgruppe kämpfen zu dürfen. Neben diesen

existieren noch die „Thrill Seekers“, die sehr aktiv offizielle Propaganda verbreiten und teilweise das Internet nur für diesen Zweck nutzen. Diese Akteure beschreiben häufig Erfahrungen, die sie vorgeben im Kampf gesammelt zu haben. Die „Moral Crusaders“ hingegen fokussieren sich auf die moralische Pflicht in Bezug auf den Kampf und dessen Belohnung in theologischer Hinsicht. Außerdem existieren noch die „Narcissists“, die politische Ursachen und persönliche Unzufriedenheit nutzen, um ein Klima der Revanche zu erzeugen, das Personen dazu verleiten soll, sich den Zielen des IS anzuschließen. Abschließend gibt es noch „Identity Seekers“, die sich bereits größtenteils im oben beschriebenen Radikalisierungsprozess befinden und auf der Suche nach eigener Identität sind. Diese Akteure werden von der vermeintlichen heroischen Männlichkeit angesprochen, die in der IS-Propaganda transportiert wird (Awan, 2017, S. 143).

Diese verschiedenen Gruppentypen beschränken sich aber keinesfalls nur auf männliche Akteure, sondern schließen Frauen ebenso mit ein. Interessant ist in diesem Zusammenhang, dass Frauen weitaus einflussreicher als Männer in den sozialen Netzwerken sind und dementsprechend prädestinierter für die Verbreitung von Propaganda sind (Huey, Inch, & Peladeau, 2019, S. 447). Analog zu den oben genannten Gruppentypen existieren solche Rollen auch innerhalb der weiblichen Anhängerschaft des IS. Dabei ist zwischen acht verschiedenen Typen zu unterscheiden, die im Folgenden kurz betrachtet werden. Die vermeintlich jüngste Gruppe unter weiblichen IS-Sympathisanten machen die „Fan-Girls“ aus, die typischerweise durch junge Frauen gebildet werden und es als „cool“ erachten, Teil eines subversiven Netzwerkes zu sein. Diese jungen Frauen schreiben über verschiedenste Themen; vom Alltag in der Schule bis hin zur Unterstützung von tödlicher Gewalt. Wenngleich sie meistens nicht selbst Propagandainhalte produzieren, so verbreiten sie offizielle IS-Propaganda zumindest. Eine andere wichtige Gruppe sind die „Baqiya-Members“, die sich selbst als wichtige unterstützende Mitglieder der

Terrorgruppe sehen und diese Unterstützung auf mehreren Ebenen leisten. Dazu gehören ideologische und emotionale, aber auch propagandistische und finanzielle Unterstützung, wenn nicht gar eine Migration in das IS-Territorium erfolgt. Neben diesen beiden Gruppen gibt es noch die „Propagandists“, die in gewisser Weise das ideologische Vollstreckungsorgan des „Islamischen Staates“ darstellen. Diese Akteure sind hochgradig religiös, was sich in ihren Propagandanachrichten widerspiegelt, in denen keinerlei persönliche Informationen zu finden sind. Sie widmen sich ausschließlich dem Ziel, Personen ideologisch zu potentiellen Rekruten zu verführen. Die „Recruiters“ sind die Gruppe, die in solchen Fällen als Kontaktperson fungiert und Informationen bereitstellt, beziehungsweise emotionale Unterstützung leistet, wenn ein Rekrut den Wunsch äußert, sich der Gruppe anzuschließen. Außerdem gibt es noch eine Gruppe namens „Muhijrat“, die ausschließlich aus Frauen besteht, die sich dem IS nicht nur ideologisch, sondern auch physisch durch eine Migration angeschlossen haben. Sie bilden in gewisser Weise das Rückgrat des sogenannten „Islamischen Staates“, indem sie nicht nur als Beispiele für eine Erfolgsgeschichte in der Propaganda dienen, sondern auch Teil der physischen Bevölkerung vor Ort sind. Den größten Teil dieser Gruppe bilden Mütter und Ehefrauen, die mit Kämpfern verheiratet sind. Neben diesen Akteuren spielen die „Widows“ ebenfalls eine zentrale Rolle, da diese einen sehr hohen Status im selbsternannten Kalifat genießen, da ihre Männer im Kampf für den IS getötet wurden und damit aus deren Perspektive für einen höheren Grund gestorben sind. Das Gegenstück zu den „Widows“ bilden „Terrorists“, die in jeglicher Weise mit terroristischen Aufgaben beauftragt sind und diese ausführen. Als letztes ist noch die Gruppe der „Leavers“ zu nennen, die sich aus unterschiedlichen Gründen vom IS getrennt haben. Dies muss nicht zwangsläufig aufgrund einer Läuterung geschehen, sondern kann auch dann der Fall sein, wenn sich stattdessen einer anderen Terrorgruppe angeschlossen wird (Huey, Inch, & Peladeau, 2019, S.

450). Ein entscheidender Unterschied, der zwischen den „Fan-Girls“ und den anderen Gruppen indes besteht, ist die Veröffentlichung des eigenen Gesichts, das ansonsten von allen anderen Rollen strikt vermieden wird. Dieser Umstand ist nicht zuletzt auf die Naivität vieler dieser jungen Frauen zurückzuführen (Ghajar-Khosravi et. al., 2016, S. 81).

Eine besondere Aufmerksamkeit verdient jedoch der Vergleich zwischen Männern und Frauen innerhalb der IS-Sympathisanten – besonders in Bezug auf das Nutzerverhalten in sozialen Netzwerken. Dabei fällt auf, dass männliche IS-Anhänger auf Twitter tendenziell mehr Follower haben als Frauen und auch öfter genannt, beziehungsweise retweetet werden. Außerdem nutzen sie viel häufiger Hashtags und Links in ihren Tweets als Frauen, wohingegen Frauen häufiger Tweets als Favoriten markieren und auch eigene Tweets häufiger als Favorit markiert werden (Pearson, 2018, S. 859f). Auch das Rollenbild, das durch Tweets teilweise ausgedrückt wird, unterscheidet sich zwischen männlichen und weiblichen IS-Sympathisanten. Während Männer dem Zweck dienen, als Kämpfer für das Kalifat ihren Beitrag zu leisten, so fällt dieser bei Frauen eher in die Kategorie Staatsbildung, folglich also Kinderreichtum und die Rolle als Ehefrau und Mutter. Diese Rollenverteilung ist auch identisch mit der, die in den offiziellen Propagandamagazinen *Dabiq* und *Rumiyah* verbreitet wird (Pearson, 2018, S. 861).

Neben diesen Unterschieden gibt es aber auch Gemeinsamkeiten, die sich vor allem auf die Reaktionen auf eine erfolgte Sperrung beziehen. So fällt ungeachtet des Geschlechts bei vielen Profilen von IS-Anhängern auf, dass diese die Anzahl der bereits erfolgten und „überlebten“ Sperrungen in die Beschreibung des eigenen Profils aufnehmen. Diese vermeintlich große Leistung gegen die „kuffar“ (Ungläubigen) spielt innerhalb der Gruppe der Sympathisanten die Rolle einer Ehrenauszeichnung (Pearson, 2018, S. 863). Im Grunde kann man sogar davon sprechen, dass eine zeitweise

Suspendierung des Profils ein so bedeutendes Ereignis im Lebenszyklus eines Twitterprofils von IS-Anhängern ist, dass eine Rückkehr schon fast wie ein Ritual zelebriert wird. Dieses Ritual vollzieht sich in drei Stufen, wobei die erste Stufe die offizielle Verkündung darstellt, dass ein Nutzer wieder da ist. Die zweite Stufe ist die virtuelle Hilfe in Form von Nennungen (@Musterbeispiel) und Retweets durch andere Nutzer und speziell angelegte „Shout-out-Accounts“, deren einziger Zweck darin besteht, Rückkehrern beim Aufbau des alten Netzwerkes zu helfen. Die dritte und finale Stufe ist dann das gegenseitige Retweeten, was häufig mit Danksagungen verbunden ist (Pearson, 2018, S. 865). Dieses Verhalten findet auch bei der Verbreitung von Propagandamaterial Verwendung, wie in Abb. 2 skizziert wurde. Um den gesamten verbreiteten Inhalt nicht vollständig zu verlieren, werden von vielen IS-Anhängern auch Back-up-Profile erstellt, die dann im Falle einer Nutzung weiterverwendet werden können (Pearson, 2018, S. 867).

Im Rahmen einer Betrachtung dieser Gruppenstrukturen wird das Ausmaß der verschiedenen Aufgaben und Gruppentypen deutlich. Interessant ist jedoch, dass sich diese Aufgaben- und Rollenverteilungen auch in den sozialen Netzwerken widerspiegeln und auf Basis des Sprachgebrauchs und Inhalts – zumindest intellektuell – eine grobe Klassifizierung möglich ist, was folglich auch Möglichkeiten für CT bereitstellt.

3 Analyse in Form korpuslinguistischer Sprachanalyse

Nachdem bislang der theoretische Unterbau behandelt wurde, wird sich im weiteren Verlauf dem experimentellen, methodischen Teil gewidmet. Dabei wird die korpuslinguistische Sprachanalyse und ihr Potential in Bezug auf die Nutzung von Textmaterial zunächst etwas genauer beleuchtet. Ebenso geht es im anschließenden Schritt um die mathematische Verifizierung der gefundenen Ergebnisse, was durch eine Signifikanzanalyse geschehen soll, um dann die SA zu betrachten, was der Einleitung in den experimentellen Teil dient. Darauffolgend wird die Datenerhebung sowie eine Erläuterung der verwendeten Textkorpora durchgeführt, um die Datenqualität und den Datenumfang möglichst genau skizzieren zu können, wobei auch die genutzten Programme und die Arbeitsschritte behandelt werden, um die Methodik nachvollziehbar zu machen. Dabei werden auch insbesondere die aufgetretenen Probleme, die sich bei der Datenerhebung und der Datenbereinigung ergeben haben, genauer betrachtet. Anschließend werden dann die eigentlichen Analysen in Form einer korpuslinguistischen Sprachanalyse und die statistischen Signifikanztests durchgeführt, die darauffolgend von einer SA inhaltlich abgerundet werden.

Diese durchgeführten Analysen stehen dabei stets im Kontext zu den Fragestellungen, die in dieser Arbeit von Belang sind und die mithilfe der Experimente beantwortet werden sollen. Zunächst geht es dabei um die Frage, ob sich Personen mit einer bestimmten Ideologie anhand ihres Sprachgebrauchs identifizieren lassen. Des Weiteren ist von Interesse, ob sich anhand korpuslinguistischer Analysen Indikatoren definieren lassen, die für eine Identifizierung solcher Personen besonders aussagekräftig sind. Zum Schluss steht außerdem noch die Frage aus, inwieweit sich die einzelnen TG in Bezug auf den Sprachgebrauch voneinander unterscheiden, beziehungsweise Gemeinsamkeiten aufweisen.

Nach der Durchführung dieser Analysen und der Präsentation der Ergebnisse werden diese im abschließenden Kapitel im Hinblick auf die oben genannten Fragestellungen diskutiert.

3.1 Ziele und Chancen einer korpuslinguistischen Analyse

Von den bereits betrachteten Verfahren sind solche besonders hilfreich, die sich auf Textstücke beziehen, die von einer Gruppe oder deren Anhängern selbst produziert wurden. Korpuslinguistische Ansätze machen sich dabei den Umstand zunutze, dass genutzte Sprache und das Weltbild einer Person unmittelbar miteinander verknüpft sind. Wörter sind demzufolge Reflektionen dessen, wie jemand die Welt sieht oder wie jemand diese zu konstruieren wünscht (Marcellino et al., 2017, S. 38). Eine besonders wertvolle Quelle ist produzierter Text auch deshalb, weil sich gleich mehrere Zugriffsmöglichkeiten ergeben. Verfasste Textstücke basieren auf der Annahme, dass jeder Autor – ähnlich einem Fingerabdruck – einen einzigartigen Schreibstil hat (Athira, & Thampi, 2015, S. 174). Geschriebener Text bietet außerdem die Möglichkeit, anhand speziell definierter linguistischer Marker Texte zu analysieren. Diese Marker werden oft auf Basis von Worthäufigkeiten und Phrasen erstellt und lassen sich im Rahmen der Nutzung von NLP mit relativ einfachen Mitteln gewinnen (Cohen et al., 2014, S. 252f). Auf vergleichbare Weise arbeiten auch Instrumente zur automatischen Textzusammenfassung sowie der Informationsextraktion, die sich auf unterschiedlichste Formen von Text anwenden lassen und in verschiedenen Aufgabenfeldern Einsatz finden können. Besonders hilfreich sind solche Instrumente bei der Erstellung von Quasi-Klassifikationen, bei der allein auf Grundlage von Textcharakteristika und Ähnlichkeitskoeffizienten ähnliche Textdokumente eingeordnet werden (Stock, & Stock, 2008, S. 372f). Gleiches gilt auch für die automatische Textzusammenfassung, bei der häufig – wie bei der Korpusanalyse – Indikatoren wie Kookkurrenzen, Satzposition

und Worthäufigkeiten eine wichtige Rolle spielen (Saggion, & Poibeau, 2013, S. 10). Auch die Extraktion von besonders wichtigen Sätzen innerhalb eines Dokumentes ist mithilfe statistischer Gewichtungen auf diese Weise möglich (Stock, & Stock, 2008, S. 399). Weitere Aufgabenstellungen in diesem Kontext sind die Eigennamenerkennung, die Identifikation von mehreren Namen für dieselbe Entität und die Extraktion von Beziehungen (Piskorski, & Yangarber, 2013, S. 26f).

Auf Grundlage dessen ist es möglich, linguistische Korpora auf Basis des Sprachgebrauchs zu erstellen, die anschließend anhand statistischer Methoden analysiert werden können und Ergebnisse liefern. Von besonderer Bedeutung sind dabei Termfrequenzen, die als Indikator für Schlüsselwörter eines Textes genutzt werden können (Saggion, & Poibeau, 2013, S. 5). Außerdem können linguistische Modelle erarbeitet werden, in denen die Frage im Vordergrund steht, welche Wörter statistisch gesehen häufiger als in anderen Texten genutzt werden, was sich auch als Grundlage für eine solche Betrachtung in anderen Korpora (wie Texten aus sozialen Netzwerken) nutzen lässt, um beispielsweise Charakteristika zu vergleichen (Marcellino et al., 2017, S. 39f). Solche Verfahren erlauben es auch, nach einzelnen semantischen Elementen zu suchen, um beispielsweise mithilfe statistischer Verfahren sprachliche Muster zu identifizieren oder diese mit anderen Eigenschaften des Textes in Bezug zu bringen und für eine automatische Inhaltsanalyse zu nutzen (Wettstein, 2014, S. 18). Besonders statistisch auffällig sind in diesem Zusammenhang Kookkurrenzen, also das gemeinsame Auftreten von Wörtern in einem fest definierten Abschnitt eines Textes, da diese in ebenjener Kombination deutlich seltener auftreten. Obwohl es immer eine potentielle Zufälligkeit für eine bestimmte Kombination gibt und manche dieser zum Beispiel einfache zusammengesetzte Ausdrücke darstellen, so spielen Kookkurrenzen dennoch in korpuslinguistischen Ansätzen eine wichtige Rolle (Marcellino et al., 2017, S. 43). Allgemein impliziert die statistische

Betrachtung eines Textes weniger ein Interesse an absoluten, als vielmehr an relativen Häufigkeiten, da nicht sprachlich normale, sondern anormale, untypische Muster entdeckt werden sollen (Grothe, 2016, S. 118).

Schwierigkeiten gestalten sich meistens nach den Mustern, die im NLP ohnehin bekannt sind und vor allem durch Synonyme („Spende“ – „Gabe“) und Koreferenzen („Anna sagt, sie liebe Tiere“) bestehen. Jedoch lassen sich solche Schwierigkeiten unter Einsatz größeren Aufwandes lösen, sofern nicht ohnehin bereits eine optimierte Softwarelösung bereitsteht (Saggion, & Poibeau, 2013, S. 5). Trotzdem sollte immer bedacht werden, dass die Komplexität natürlicher Sprache nicht zuletzt aufgrund ihres Gebrauchs von Metaphern, Ironie, satzübergreifenden Rückbezügen und anderen Konstruktionen nach wie vor eine Herausforderung für NLP darstellt (Wettstein, 2014, S. 24f).

Etwas einfacher hingegen verhält es sich mit textstatistischen Verfahren, die ebenfalls Teil einer korpuslinguistischen Analyse sind. Diese beruhen in erster Linie auf der einfachen Aufzählung von Text-, Satz- und Wortstatistiken, die vor allem mithilfe eines Computers extrem schnell und genau vollzogen werden können und damit einen Vorsprung in Bezug auf Zeit und Genauigkeit im Verhältnis zu manuellem Zählen aufstellen. Durch diese einfach durchzuführenden Verfahren lassen sich innerhalb kürzester Zeit Worthäufigkeitslisten erstellen, die beispielsweise einen schnellen Überblick über Thematiken geben können oder zur Erstellung von Stoppwortlisten genutzt werden können (Scharkow, 2012, S. 61f). Durch die stetig wachsende Zahl von Indikatoren, die zu korpuslinguistischen Analysen herangezogen werden, ergeben sich des Weiteren immer neue Anwendungsfelder, die von der Zuordnung von Shakespeare-Stücken über Autorenidentifikation (Authorship Analysis) bis hin zum modernen OM reichen (Scharkow, 2012, S. 64). Diese gemeinsamen Vorkommen von Wörtern lassen sich anschließend in eine Kontingenztafel oder auch

Ähnlichkeitsmatrix überführen, um anschließend weitere statistische Verfahren anwenden zu können (Scharkow, 2012, S. 66). Solche textbasierten korpuslinguistischen Analysen eignen sich grundsätzlich hervorragend für sämtliche Textarten, vereinen sowohl linguistische, als auch statistische Indikatoren und sind in vielerlei Hinsicht anwendbar und interpretierbar, dabei jedoch thematisch völlig unabhängig. Aus diesem Grund ist es möglich, für gezielte Fragestellungen ein sehr passgenaues Studiendesign zu entwerfen, anhand dessen nach unterschiedlichen Indikatoren verfahren werden kann.

3.2 Weiterführende Methoden und Strategien zur Ergebnisverifizierung

Korpuslinguistische Analysen haben das Potential, aussagekräftige Ergebnisse in Bezug auf die Wortwahl und Charakteristik der verwendeten Sprache zu liefern. Da jedoch allein das Auftreten von bestimmten Kookkurrenzen oder Termen ohne Kontext wenig aussagekräftig ist, steht die Frage im Raum, inwiefern diese als charakteristisch für eine gruppenspezifische Sprache definiert werden können.

Um diese Charakteristik mathematisch beweisen zu können, bietet sich ein Signifikanztest an, der sich der Frage widmet, ob mit ausreichend großer Wahrscheinlichkeit angenommen werden kann, dass der Frequenzunterschied von einzelnen Termen innerhalb der Korpora nicht zufällig zustande gekommen ist. Dafür werden für häufig verwendete Terme Kontingenztabelle erstellt, die dann im nächsten Schritt normalisiert und durch einen Signifikanztest in Form eines Chi-Quadrat-Tests untersucht werden. Auf diese Weise lässt sich – je nach gewünschter Signifikanz – mit ausreichend großer Wahrscheinlichkeit belegen, dass der Unterschied der Frequenzen von Termen zwischen zwei Testkorpora nicht zufällig erfolgt ist. Im Kontext der Sprachanalyse von Terrorgruppen lässt sich somit im Idealfall nachweisen, dass eine signifikante Korrelation zwischen einer Termfrequenz

und einer bestimmten Gruppe existiert, was letztlich durch ideologische Aspekte begründet ist.

Die mathematische Formel des Chi-Quadrat-Tests definiert sich folgendermaßen:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Die Variable O beschreibt in dieser Formel die beobachtete (observed) Frequenz eines Terms innerhalb eines Korpus, während E eine Variable für die erwartete (expected) Frequenz darstellt. Anhand einer Tabelle für kritische Werte und unter Berücksichtigung des korrekten Freiheitsgrades, der abhängig von der Dimension der Kontingenztafel ist, lässt sich die mathematische Signifikanz ermitteln, die entscheidend für die Beantwortung der Frage ist, ob die Nullhypothese mit ausreichender Wahrscheinlichkeit verworfen werden kann und die eigene Hypothese zutrifft (Athira, & Thampi, 2015, S. 182).

Neben dem statistisch betrachteten Sprachgebrauch existiert auch die Möglichkeit, mithilfe einer SA den verwendeten Sprachgebrauch zu untersuchen. Ein großer Vorteil der SA ist die automatische Auswertung von Material im Hinblick auf Meinung und Haltung. Dabei werden alle Ausdrücke in textlicher Sprache untersucht und anschließend in Werten skaliert, um Vergleiche ziehen zu können. Im Endeffekt kann textliches Material aus einer Vielzahl von Quellen durch solche Programme in „positiv“, „neutral“ oder „negativ“ klassifiziert werden. Dies geschieht in der Regel je nach Gebrauch der Wörter und dem Kontext. Dabei zeigt sich die SA als sehr zeit-, kosten- und arbeitseffizient und als besonders ergiebig für Textmaterial aus Twitter (Macnair, & Frank, 2018b, S. 443f).

Durch die Kombination verschiedener Verfahren kann sich ein multidimensional abgebildetes Ergebnis ergeben, was eine Verbesserung der Ergebnisse mit sich bringt. Nachdem zunächst verschiedene Analysen im

Rahmen der korpuslinguistischen Untersuchung durchgeführt werden, sollen im anschließenden Schritt einige der Ergebnisse mathematisch auf Signifikanz hin untersucht werden. Die SA im letzten Schritt des methodischen Teils bildet im Idealfall ein Ergebnis, das die bereits erzielten Ergebnisse stützt.

3.3 Datenerhebung und Korpusaufbau

Die Erhebung von Daten ist für jede Art von Forschung essentiell, um eine empirische Grundlage für weiterführende Analysen zu haben. In konkretem Bezug zu korpuslinguistischen Analysen bestehen diese Daten aus textuellem Datenmaterial. Um für die vorliegende Arbeit relevante Daten im Kontext der Fragestellungen erheben zu können, musste im ersten Schritt entschieden werden, welche Medien überhaupt nach relevantem Textmaterial durchsucht und welche Daten genau extrahiert werden sollen. Aufgrund der Tatsache, dass der erste Korpus (Korpus A) als Vergleichskorpus dienen soll und den Sprachgebrauch der Terrorgruppe IS reflektieren soll, besteht dieser Korpus ausschließlich aus Textmaterial, welches aus den oben betrachteten offiziellen Propagandamagazinen *Dabiq* und *Rumiyah* extrahiert wurde (siehe Anhang 22). Diese Magazine wurden von der Homepage der non-profit Organisation „Clarion Project“ gewonnen, die diese Propagandamagazine in englischer Sprache für die Forschung und zu Bildungszwecken öffentlich zugänglich macht. Nach dem Download der Magazine im PDF-Format und der anschließenden Textextraktion in ein gängiges .txt-Format war der Korpus A bereit für die Datenbereinigung, wenngleich die graphischen Abbildungen eine Verarbeitungshürde darstellten, da manche Textkodierungen nicht automatisch erkannt wurden und daher manuell bearbeitet werden mussten, um keinen Informationsverlust erleiden zu müssen (Hirschmann, 2019, S. 22). Deutlich schwieriger gestaltete sich die Datenerhebung für den Korpus B, in dem Textmaterial in Form von Tweets von offiziellen IS-Kanälen, aber auch

von deren Anhängern und Sympathisanten gesammelt werden sollte. Aufgrund der oben bereits beschriebenen Tatsache, dass Tweets, beziehungsweise Profile mit IS-Bezug seit einigen Jahren konsequent gelöscht werden, gestaltete sich die Suche nach relevanten Tweets, die über die Twitter API extrahiert werden können, als schwierig. Durch die große und seit Jahren täglich wachsende Datenbank von Prof. Dr. Robert Jäschke wurde jedoch die Möglichkeit geboten, auf ebenjene bereits gelöschten Tweets in Form von Kopien zugreifen zu können und für eine Analyse im Rahmen dieser Arbeit gewinnen zu können. Dafür wurde ein Zugang auf dem Server mit der Datensammlung freigeschaltet, über den sich via VNP Zutritt verschafft werden konnte. Die Twitterdaten lagen auf diesem Server in großen Dateien in komprimiertem JSON-Format vor, die aus zeitökonomischen Gründen am besten mit einem Analysetool verarbeitet werden konnten, das die Dateien nicht entpackt und erst anschließend ausliest, sondern dies in komprimiertem Zustand leisten konnte. Aus diesem Grund wurde das Tool „Apache Spark 2.4.5“ genutzt, das eine Analyse-Engine für Datenverarbeitung in großem Maßstab ist. Dieses Tool basiert auf der Programmiersprache *Scala*, die eigens für den Zweck der Datenerhebung für diese Arbeit erlernt wurde. Nachdem ein passendes *Scala*-Script geschrieben wurde (siehe Anhang 1) und die ersten Suchanfragen positiv ausfielen, begann die eigentliche Datenerhebung.

Aufgrund der Größe der Dateien belief sich eine einzelne Suchanfrage auf circa 25 Minuten pro Jahreszahl, die durchsucht wurde. Bei insgesamt 105 verschiedenen Suchanfragen pro Jahr und einem betrachteten Analysezeitraum von sechs Jahren ergeben sich so 630 Suchanfragen zu je 25 Minuten, was sich am Ende auf eine – wohlgernekt optimal gerechnete – Gesamtdauer der Datenerhebung für Korpus B von 15 750 Minuten oder knapp 11 Tage beläuft. Die Suchanfrageterme bestanden dabei zu großen Teilen sowohl aus Hashtags, die von der Terrorgruppe im Rahmen ihrer

Online-Strategie genutzt wurden, als auch aus Profilen, die bereits in anderen Forschungsarbeiten als Profile mit IS-Bezug definiert wurden (siehe Anhang 2). Diese Profile dienten dann auch als Ausgangspunkt, um nach dem Schneeballprinzip über Profilinformationen weitere relevante Profile zu finden und deren Inhalt in die Analyse miteinzubeziehen, was sich auch in anderen Arbeiten als sehr effektiv erwiesen hat (Klausen, Marks, & Zaman, 2016, S. 5f). Alle auf diese Weise extrahierten Ergebnisse wurden anschließend ebenfalls in gängigem .txt-Format gespeichert und waren für die Datenbereinigung bereit.

Korpus C hingegen, der aus aktuellen Twitterdaten bestehen sollte, die sowohl von Muslimen, als auch von Nicht-Muslimen verfasst werden sollten, war deutlich einfacher zu erstellen. Zunächst stellte sich jedoch die Frage, nach welchen Kriterien überhaupt Tweets gesammelt werden sollten, um einen vergleichbaren Korpus zu haben, der angemessen für die Fragestellungen dieser Arbeit genutzt werden kann. Im Endeffekt wurde sich auf einzelne Hashtags festgelegt, die schon bei der Erstellung des Korpus B gute Ergebnisse lieferten, und von denen anhand des Schneeballprinzips auf weitere Hashtags und Profile geschlossen werden konnte (siehe Anhang 3). Aufgrund der Tatsache, dass der Korpus sowohl die Sprache von Muslimen, als auch von Nicht-Muslimen abbilden soll, wurde auf verschiedene Suchfilter wie Geo-Tag oder andere profilbezogene Filter verzichtet, um nicht beispielsweise ausschließlich Tweets aus einem mehrheitlich muslimischen oder nichtmuslimischen Land zu extrahieren.

Um die Tweets über die Twitter API zu extrahieren, wurde das Tool „Chorus Tweetcatcher Desktop Edition“ genutzt, für dessen Nutzung lediglich ein Twitter-Developer-Account und eine eigene Applikation erstellt werden musste. Für die Datenerhebung wurde im Zeitraum einer Woche (03.04.2020 – 10.04.2020) jeden Tag 30 Suchanfragen abgesetzt und die Ergebnisse im .txt-Format für die weitere Datenbereinigung vorbereitet. Im Ergebnis steht

eine relativ vielseitige Mischung aus Tweets, die thematisch verwandt mit den Suchtermen von Korpus B sind und damit keine künstliche thematische Distanz zwischen den beiden Korpora erzeugen. Dies schließt jedoch keineswegs aus, dass sich in Korpus C auch solche Textmaterialien befinden, die sich ausgesprochen kritisch mit dem Islam beschäftigen oder diesen gar beleidigen.

Im Endergebnis entstanden also drei verschiedene Korpora, die inhaltlich folgendermaßen aufgestellt sind: Korpus A beruht ausschließlich auf Textmaterial, dass hochoffiziell von der Terrorgruppe IS in Form der Propagandamagazine veröffentlicht wurde; Korpus B beruht auf Textmaterial in Form von Tweets, die sowohl von offiziellen IS-Akteuren, als auch von Sympathisanten veröffentlicht wurden; Korpus C hingegen besteht aus Textmaterial in Form von Tweets, die zwar thematisch verwandt, aber sehr breit extrahiert wurden, um ein realeres Abbild der Gesellschaft, aber dennoch keine künstliche thematische Distanz zu erzielen.

Wichtig in diesem Zusammenhang sind noch einige Feinheiten, die die Charakteristik der Korpora betreffen. So wurde bei der Extraktion der Tweets darauf geachtet, dass größtenteils nur englischsprachige Tweets aufgenommen werden, obwohl arabischsprachige noch viel häufiger zu finden waren und sich auch in der Formulierung noch von den englischen Tweets unterscheiden dürften. Durch die Tatsache, dass die meisten verwendeten Tools auf englischer Sprache basieren und in dieser am meisten elaboriert sind, auf der anderen Seite aber auch die für die Korpusanalyse wichtigen Propagandamagazine als Vergleichsbasis in englischer Sprache vorliegen, besteht durch die rein englischfokussierte Auswahl kein nennenswerter Nachteil in Bezug auf die Analyseergebnisse. Übersetzungen der arabischsprachigen Tweets haben sich im Vorfeld dieser Arbeit in Form von Tests jedenfalls als sehr unzureichend erwiesen, weshalb auf diese verzichtet wurde. Des Weiteren finden sich in den Korpora B und C auch teilweise

Tweets, die thematisch nichts mit dem Islam oder der Terrorgruppe zu tun haben. Dabei handelt es sich meist um Nutzer, die beispielsweise die oben bereits behandelten „Fan-Girls“ darstellen oder Terme wie „Jihad“ als Hyperbel in anderem Kontext nutzen. Aufgrund der Tatsache, dass auch selbst auf offiziellen IS-Profilen über Themen geschrieben wurde, die inhaltlich nicht in Bezug zum Islamismus stehen (beispielsweise IS-Kämpfer mit Katze) und in den Magazinen auch das einfache Leben im Kalifat beschrieben wird, wurden diese vermeintlich „nicht-radikalen“ Tweets ebenfalls in die Korpora mitaufgenommen, da auch diese Teil der Propagandastrategie des IS sind.

Nach der Sammlung der Rohdaten und der Einteilung des Materials in den jeweiligen Korpus begann die Datenbereinigung, die ebenfalls essentiell für jede Analyse ist. Klassische Verfahren zur Bereinigung des Textmaterials sind insbesondere für extrahierte Daten aus sozialen Medien wichtig, da diese sehr häufig „Rauschen“, beziehungsweise wenig aussagekräftigen Ballast beinhalten. Dazu gehören neben Hyperlinks und Referenzen zu anderen Nutzern auch Hashtags und die Abkürzung RT für Retweet, wenn es sich um Material von der Plattform Twitter handelt. Neben diesen Eigenschaften müssen auch sich wiederholende Buchstaben (beispielsweise „nooooooooo“ für „no“) adäquat ersetzt werden. Um eine Normalisierung des gesammelten Materials zu erreichen, sollten außerdem alle Wörter durchgehend klein geschrieben sein, damit Wörter wie „Jihad“ und „jihad“ zusammengefasst werden und die Analyseergebnisse nicht negativ beeinflussen. Die Kombinationen von Wörtern, die sich aufgrund von Dialekten („color“ und „colour“) unterscheiden, sollten im Idealfall ebenfalls zusammengeführt werden (Macnair, & Frank, 2018b, S. 445f). Die anderen klassischen NLP-Verfahren (Synonym- und Homonymanpassung, Anpassung der Pluralformen, Lemmatisierung und Stemming) können je nach Bedarf angewendet werden, sind aber je nach Zielsetzung nicht zwingend notwendig

und im Rahmen korpuslinguistischer Arbeiten auch nicht unumstritten, da unter der Vorbehandlung dieser Textdaten die sprachliche Charakteristik und Vielfalt der Texte leiden könnte. Dieses sogenannte Preprocessing sollte daher mit Umsicht vorgenommen werden, um die Datenbasis schonend und sinnvoll aufzubereiten und keine unerwünschten Veränderungen vorzunehmen (Scharkow, 2012, S. 52).

Im Rahmen dieser Arbeit wurde für die drei Korpora ein äußerst schonendes Preprocessing gewählt, um die gruppenspezifische Sprache, die im Fokus steht, nicht unnötig zu verändern. Dies wurde ausschließlich über Regular Expressions (RE) erreicht, da die Menge an Daten intellektuell nicht von einer Person unter zeitökonomischen Gesichtspunkten bereinigt werden kann. Diese Vorgehensweise gleicht einem algorithmischen Ansatz, der tendenziell weniger aufwendig und sprachunabhängiger als ein linguistischer Ansatz ist (Scharkow, 2012, S. 113). Für den Korpus A, der aus Textmaterial besteht, welches aus den Propagandamagazinen extrahiert wurde, wurde außer redundanten Informationen wie Seitenzahlen und arabischer Schrift wenig entfernt, um die gruppenspezifische Sprache nicht zu verfälschen. Vielmehr wurden Doppelabsätze und -leerzeichen, fehlende Leerzeichen nach Satzenden und Formatierungsfehler korrigiert sowie das gesamte Material in Kleinbuchstaben geändert. Auf diese Weise wurden die Textdaten bestmöglich für das Korpusanalyseprogramm vorbereitet, ohne dabei einen Informationsverlust im Hinblick auf Sprache oder Inhalt zu provozieren. Ähnlich wurde auch mit den Korpora B und C verfahren, die aus Textmaterial bestehen, welches von Twitter extrahiert wurde. Jeweils wurden mithilfe von RE unwichtige Informationen wie Hyperlinks, Profilreferenzen, Abkürzungen für Retweets und Buchstabenwiederholungen entfernt, Leerzeichen- und Absatzfehler korrigiert und abschließend das ganze Material auf Kleinbuchstaben umformatiert, damit ordentlich strukturierte Dateien entstehen, die in die Korpusanalyseprogramme eingespeist werden können

(siehe Anhang 4). Alle weiteren Informationen zu Profilen, die sich von Nutzernamen über Freundesanzahl bis hin zu weiteren visuellen Charakteristiken erstrecken, sind zwar für eine Netzwerkanalyse sinnvoll, im Rahmen dieser Arbeit aber weniger von Bedeutung und wurden daher auch entfernt. Dennoch handelt es sich dabei selbstverständlich auch um sehr reichhaltige Informationen, die im Rahmen von anderen Fragestellungen und Analysen genutzt werden können.

Nachdem abschließend alle für die korpuslinguistische Sprachanalyse benötigten Daten erhoben und bereinigt wurden und in für das Programm passende Datenformate umgewandelt wurden, wurde sich der eigentlichen Analyse zugewendet.

3.4 Sprachanalyse und Anwendung der Indikatoren

Nach der Erhebung und Aufbereitung der Korpora wird sich der eigentlichen korpuslinguistischen Sprachanalyse gewidmet. Dabei steht zunächst die Frage im Vordergrund, ob sich anhand des Sprachgebrauchs Personen identifizieren lassen, die einer konkreten Ideologie nahestehen. Dies soll im Hinblick auf Ähnlichkeiten innerhalb des Sprachgebrauchs geschehen, indem Korpus A, der auf offiziellem IS-Textmaterial beruht, als Referenzmodell dient, um eventuelle Ähnlichkeiten oder Unterschiede zu diesem in den Korpora B (IS-bezogene Inhalte von Akteuren und deren Sympathisanten auf Twitter) und C (thematisch ähnliche, aber IS-unabhängige Nutzer) nachzuweisen. Anhand dieses Verfahrens soll ebenso die Frage betrachtet werden, welche Indikatoren sich als besonders effektiv für eine Identifizierung dieser Personen erweisen und welche Indikatoren dafür weniger geeignet sind. Abschließend wird sich noch der Frage gewidmet, welche Unterschiede und Gemeinsamkeiten im Sprachgebrauch zwischen den jeweiligen TG bestehen, um eventuelle Rückschlüsse auf die Gruppen ziehen zu können.

Für die Durchführung dieser Analyse wurden die Programme „AntConc“ und „CorpusExplorer v2.0“ ausgewählt, da diese neben umfangreichen Funktionen besonders intuitiv und nutzerfreundlich zu bedienen sind und sich außerdem bei Pretests als am zuverlässigsten erwiesen haben. Nach der Einspeisung der drei Korpora lassen sich folgende Aufteilungen der Dokumente und Token erkennen (Tab. 1).

	Korpus A	Korpus B	Korpus C	Gesamt
Dokumente	28	174	30	232
Token	708 552	186 369	4 126 656	5 021 577

Tabelle 1: *Dokument- und Tokenverteilung der jeweiligen Korpora* (Eigene Darstellung)

Auf diese drei verschiedenen Korpora sollen nun im Folgenden verschiedene Verfahren mithilfe der beiden Programme angewendet werden, um bestimmte Indikatoren zu erzeugen, anhand derer sich Charakteristiken des Sprachgebrauchs feststellen lassen. Um einige Standardwerte zu erhalten, die Grundvoraussetzungen für diese Verfahren sind, werden mithilfe des Programms „AntConc“ Eigenschaften der jeweiligen Korpora ermittelt. Gegenstand der Betrachtung sind dabei zunächst die 20 häufigsten Terme in jedem Korpus (Tab. 2), allerdings bereinigt um eine eigens angelegte korpuspezifische Stoppwortliste (siehe Anhang 4), um Terme mit wenig Aussagekraft zu entfernen. Diese nachfolgende Tabelle dient im weiteren Verlauf der Arbeit als Basis für alle weiteren Berechnungen.

Rang	Korpus A	Korpus B	Korpus C
1	allah (7674)	isis (2167)	allah (64634)
2	state (2164)	jihad (1916)	isis (50537)
3	islamic (2093)	islamicstate (1447)	islamic (17806)
4	people (1972)	islam (1377)	quran (16085)

5	muslims (1472)	caliphate (914)	state (14588)
6	islam (1190)	myjihad (898)	muslim (13879)
7	religion (1184)	muslim (871)	islam (12117)
8	messenger (1099)	syria (868)	people (10084)
9	soldiers (1028)	muslims (649)	good (9021)
10	muslim (1024)	islamic (635)	love (7831)
11	prophet (927)	iraq (604)	time (7479)
12	abu (826)	khilafah (568)	life (6106)
13	abū (810)	isil (547)	muslims (6061)
14	war (796)	allah (490)	world (5371)
15	kill (789)	sharia (424)	forgiveness (5307)
16	report (758)	state (424)	attack (5069)
17	fight (648)	terrorism (362)	syria (4897)
18	lord (644)	people (349)	caliphate (4868)
19	jihad (639)	daesh (337)	day (4735)
20	day (624)	kill (324)	iraq (4712)

Tabelle 2: *Top-20 Terme (+ absolute Häufigkeit) in allen Korpora* (Eigene Darstellung)

Da aufgrund der unterschiedlichen Größe der Korpora eine reine Betrachtung der absoluten Vorkommen der Terme keine adäquate Variante ist, um die Terme innerhalb der verschiedenen Korpora miteinander zu vergleichen, sollte eine Betrachtung der relativen Häufigkeit der Terme vorgenommen werden, die sich leicht anhand der Daten aus Tabelle 1 und Tabelle 2 herleiten lässt (Tab. 3). Diese Normalisierung dient dem Zweck, bei einer statistischen Auswertung die vorhandenen Werte miteinander vergleichbar zu machen (Hirschmann, 2019, S. 187). Dies ist im vorliegenden Fall notwendig, da

aufgrund der unterschiedlichen Korpusgrößen kein Vergleich anhand absoluter Termfrequenzen vorgenommen werden sollte.

Rang	Korpus A	Korpus B	Korpus C
1	allah (0,1083)	isis (0,1163)	allah (0,1566)
2	state (0,0305)	jihad (0,1028)	isis (0,1225)
3	islamic (0,029)	islamicstate (0,0776)	islamic (0,0431)
4	people (0,0278)	islam (0,0739)	quran (0,0390)
5	muslims (0,0208)	caliphate (0,0490)	state (0,0354)
6	islam (0,0168)	myjihad (0,0482)	muslim (0,0336)
7	religion (0,0167)	muslim (0,0467)	islam (0,0294)
8	messenger (0,0155)	syria (0,0466)	people (0,0244)
9	soldiers (0,0145)	muslims (0,0348)	good (0,0219)
10	muslim (0,0145)	islamic (0,0341)	love (0,0190)
11	prophet (0,0131)	iraq (0,0324)	time (0,0181)
12	abu (0,0117)	khilafah (0,0305)	life (0,0148)
13	abū (0,0114)	isil (0,0294)	muslims (0,0147)
14	war (0,0112)	allah (0,0263)	world (0,0130)
15	kill (0,0111)	sharia (0,0228)	forgiveness (0,0129)
16	report (0,0107)	state (0,0228)	attack (0,0123)
17	fight (0,0091)	terrorism (0,0194)	syria (0,0119)
18	lord (0,0091)	people (0,0187)	caliphate (0,0118)
19	jihad (0,0090)	daesh (0,0181)	day (0,0115)
20	day (0,0088)	kill (0,0174)	iraq (0,0114)

Tabelle 3: *Top-20 Terme (+ relative Häufigkeit in Promille) in allen Korpora*
(Eigene Darstellung)

Allein anhand des Rankings in Tabelle 3 lassen sich bereits Gemeinsamkeiten bei der Nutzung der einzelnen Terme erkennen. Insgesamt sieben Terme sind

in allen drei Korpora in den Top-20 Termen enthalten („allah“, „state“, „islamic“, „people“, „muslims“, „islam“, „muslim“), wenngleich die relativen Häufigkeiten dieser Terme unterschiedlich ausgeprägt sind, wie sich in folgendem Diagramm (Abb. 4) eindeutig erkennen lässt.

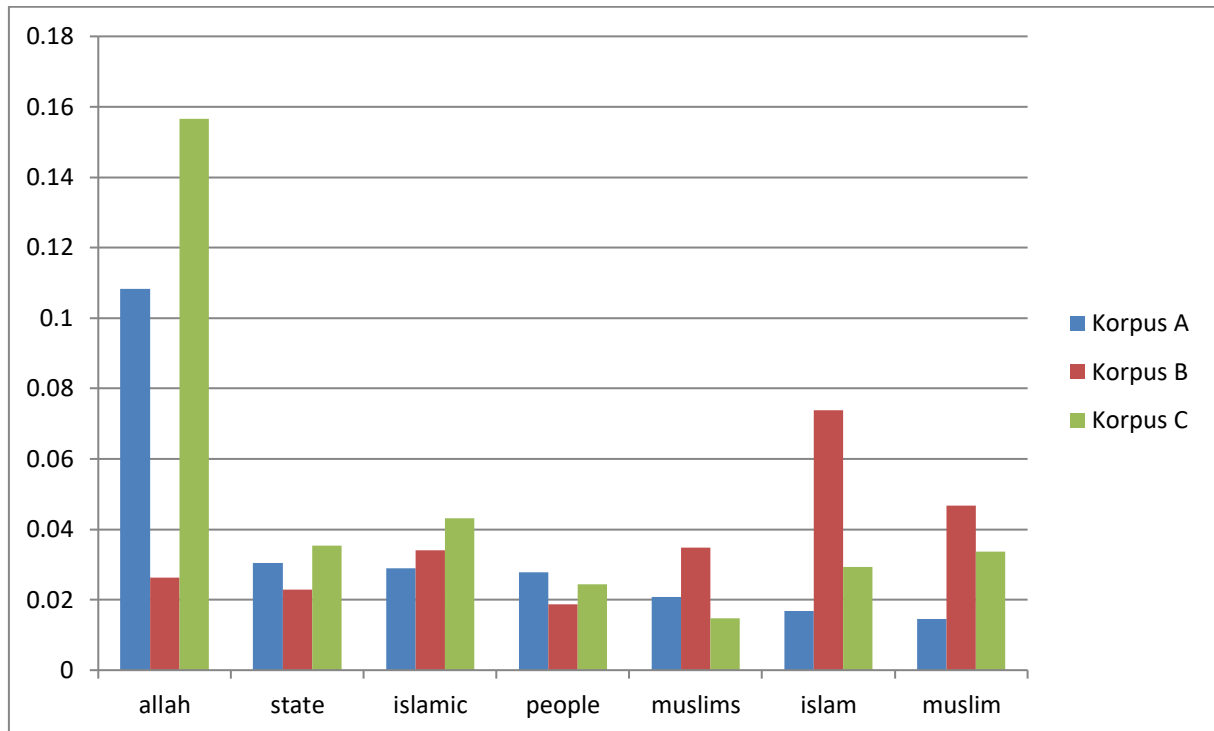


Abbildung 4: *Vergleich der relativen Häufigkeiten von sieben Termen in Promille* (Eigene Darstellung)

Anhand der Abbildung lassen sich sowohl Ähnlichkeiten, als auch Unterschiede in Bezug auf die Verwendung einzelner Terme erfassen. Dabei fällt vor allem auf, dass zwischen dem Referenzkorpus A (IS-Magazine) und Korpus B (IS-Akteure auf Twitter) keine durchgehend ähnliche Verwendung von Termen stattfindet, obwohl dies naheliegend zu sein scheint. Auf der anderen Seite gibt es aber auch Ähnlichkeiten, wie sich an den Termen „state“, „islamic“, „people“ und „muslims“ zeigt. Der Vergleich zwischen dem Referenzkorpus A und dem Korpus C (verschiedene Twitternutzer) zeigt allerdings im Mittel betrachtet eine größere Nähe zueinander. Viel aussagekräftiger werden die relativen Häufigkeiten jedoch, wenn in jedem Korpus noch andere vorkommende Terme betrachtet werden. So fällt auf,

dass im Korpus A viele Terme mit theologischem Bezug verwendet werden („religion“, „messenger“, „prophet“, „lord“), was in Kombination mit den martialischen Termen („soldiers“, „war“, „kill“, „fight“, „jihad“) für die Theorien spricht, die sich bereits ausführlich mit der Propagandastrategie des IS beschäftigt haben (siehe Kapitel 2.3.1). In Korpus B finden sich vor allem viele Terme, die die Gruppe in irgendeiner Weise bezeichnen („isis“, „isil“, „islamicstate“, „daesh“) oder sich mit dem selbsternannten Kalifat beschäftigen („caliphate“, „khilafah“ [als arabisches Wort für Kalifat], „sharia“ [im Sinne des islamischen Regelwerkes]). Dies macht angesichts der Tatsache, dass die sozialen Netzwerke von IS-Akteuren und deren Sympathisanten vereinnahmt werden, um staatsbildende Propaganda zu verbreiten, sogar Sinn. Korpus C hingegen fällt vor allem dadurch auf, dass viele Terme auftreten, die besonders positiv („good“, „love“, „forgiveness“) oder zumindest neutral („time“, „life“, „world“, „day“) konnotiert sind, obwohl keine direkte semantische Beziehung zu IS-relevanten Themen vorhanden zu sein scheint. Die Betrachtung einer breit veranlagten TG, die sich zu verwandten Themen äußert, wie es in Korpus C beabsichtigt worden war, schlägt sich also in der relativen Häufigkeit der verwendeten Terme nieder, wenngleich die alleinige Betrachtung dieser Werte keine auffälligen Erkenntnisse hervorbringt. Vielmehr lässt sich auf Basis dieser Werte ein erstes Bild des Korpusmaterials zeichnen.

Interessant ist außerdem die Betrachtung der relativen Häufigkeit einzelner Terme in Bezug auf die Verwendung über die Zeit, die sich mithilfe eines Konkordanz-Plots zeichnen lässt. Anhand dieser Funktion lässt sich die Verwendung eines Terms innerhalb eines Dokumentes graphisch darstellen, was umso aussagekräftiger wird, wenn gleich mehrere Dokumente betrachtet und miteinander verglichen werden und was als Grundlage für die Wichtigkeit eines Terms aus Sicht des Verfassers betrachtet werden kann (siehe Anhang 5 und 6). Im folgenden Beispiel wurde der häufigste

theologische aussagekräftige Term aus den IS-Propagandamagazinen („allah“) ausgewählt, die relative Häufigkeit des Vorkommens in jeder Ausgabe berechnet und dies über jede einzelne erschienene Ausgabe geplottet, was auf diese Weise die Verwendung dieses Terms im Rahmen einer Zeitachse darstellt. Zum Vergleich wurde analog mit einem neutralen Term („state“) verfahren (Abb.5).

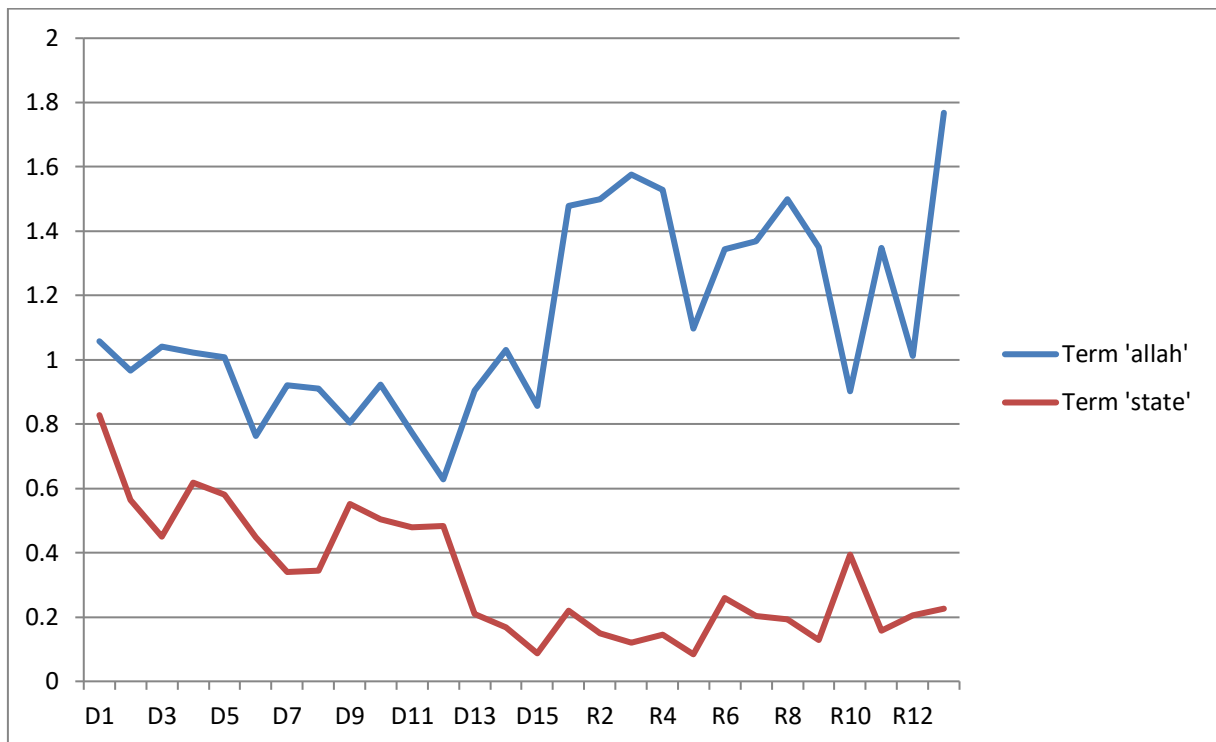


Abbildung 5: *Prozentuale relative Termhäufigkeit verteilt nach Magazinen* (Eigene Darstellung)

Anhand dieser Darstellung erkennt man, dass die relative Verwendung des Terms „allah“ sich vor allem mit Erscheinung des Magazins *Rumiyah* deutlich erhöht hat. Daraus lässt sich schließen, dass die theologische Komponente in *Rumiyah* eine deutlich größere Rolle spielt, als es noch in *Dabiq* der Fall war. Analog dazu lässt sich in Bezug auf den Term „state“ schlussfolgern, dass dieser in *Dabiq* eine deutlich wichtigere Position eingenommen hat, als es seit Erscheinung von *Rumiyah* der Fall gewesen ist. Der statistische Ausreißer in R10 ließe sich eventuell darauf zurückführen, dass der militärische Druck auf

den IS zum Zeitpunkt der Erscheinung so massiv war, dass die propagandistische Nutzung des Wortes „state“ eine Art Trotzreaktion provozieren sollte, was so auch schon in anderen Arbeiten beschrieben wurde (siehe Kapitel 2.3.1).

Weitere interessante Erkenntnisse liefert die Betrachtung von Clustern, die entweder termbezogen oder auf Basis von n-Grammen durchgeführt werden können und als Vergleichsgrundlage dienen, um die Häufigkeit von konkreten Wortsequenzen zu untersuchen, beziehungsweise deren semantisches Umfeld zu betrachten. Um die Vergleichbarkeit zwischen allen drei Korpora zu gewährleisten, wurde für die termbezogene Clusteranalyse im Folgenden der Term „allah“ betrachtet, da dieser in allen Korpora im Rahmen der Top-20-Liste und mit entsprechender Häufigkeit aufgetreten ist (Tab. 2). Dafür wurden die Suchparameter so eingestellt, dass die Clustergröße zwischen drei und fünf liegt, da kleinere Cluster wenig aussagekräftig sind und größere Cluster nicht aufgetreten sind. Außerdem wurde immer eine passende Minimalfrequenz vorausgesetzt, um unwichtige Wortkombinationen herauszufiltern. Diese sah bei Korpus A 50 Vorkommen, bei Korpus B sechs und bei Korpus C 600 Vorkommen vor, sodass in etwa die ersten 50 bis 70 Sequenzen übersichtlich miteinander verglichen werden können. Nach der Einspeisung aller drei Korpora ergaben sich Wortsequenzen (siehe Anhang 7), die sich teils zwar unterscheiden, aber klar den Kontext verdeutlichen, in welchem der benutzte Term „allah“ steht. Auf diese Weise ist es möglich, mithilfe von Clusteranalysen einen Term in einen konkreten Kontext einzuordnen und dessen Verwendung in Kombination mit anderen Wörtern zu untersuchen. In der zweiten Clusteranalyse wurde die termunabhängige n-Gramm-Clusteranalyse durchgeführt, die sich vor allem dadurch auszeichnet, dass gebräuchliche Ausdrücke unabhängig von spezifischen Termen berechnet werden. Auch in dieser Analyse wurden die Parameter zwischen drei und fünf festgelegt, jedoch ergänzt um Minimalfrequenzen, die sich für

Korpus A auf 130, für Korpus B auf 125 und für Korpus C auf 3050 beliefen (siehe Anhang 8).

In den extrahierten Wortsequenzen, die termbezogen auf „allah“ durchgeführt wurden, ist besonders auffällig, dass – bezogen auf den Referenzkorpus A – in Korpus B, der die genutzte Sprache von IS-Akteuren und deren Sympathisanten abbildet, große Gemeinsamkeiten herrschen, wie beispielsweise die Verwendung von Phrasen zeigt, die die Größe und Mächtigkeit der islamischen Gottheit demonstrieren sollen („messenger“, „permission“, „grant victory“, „sake of allah“, „lord“, „cause of allah“). Wenngleich die Sequenzen nicht immer identisch sind, so lässt sich doch feststellen, dass eine kontextspezifische Nähe existiert, da in beiden Korpora ehrfürchtig von „allah“ geschrieben wird, was sich auch dadurch ausdrückt, dass von ihm in der dritten Person geschrieben wird und dieser als oberste Institution sprachlich dargestellt wird. In Korpus C, der die Sprache verschiedener Nutzer abbildet, fällt hingegen auf, dass häufiger in der zweiten Person von ihm geschrieben wird und eher Wünsche an ihn gerichtet werden („allah, even if you“, „dear allah, please make“, „allah, forgive me“). Dies zeigt in deutlicher Weise, dass eine viel größere Gemeinsamkeit zwischen Korpus A und B, als zwischen Korpus A und C existiert. Bei der terminabhängigen Clusteranalyse, die auf n-Grammen beruht, ergibt sich ein ähnliches Bild, wenngleich die Sequenzen anders ausfallen. Während im Referenzkorpus A (also den IS-Propagandamagazinen) neben dem Gebrauch arabischer Wörter besonders die Akteure und das selbsternannte Kalifat im Vordergrund stehen („the islamic state“, „soldiers of the“, „of the khilafah“), steht in Korpus B vor allem der Jihad und die Radikalisierung im Vordergrund („so what is jihad“, „solution to the radicalisation“). Interessant ist außerdem das häufige Auftreten der Phrase „jihad watch this video“ in Korpus B, die häufig in Kombination mit Links genutzt wurde, die zu Propagandavideos des IS führten. Diese Ergebnisse zeigen also den direkt

Bezug von Twitterdaten und massenweisen Posts mit jihadistischen Inhalten und der IS-Propagandastrategie, wenngleich die staatsbildende Propaganda sich anhand der Twitterdaten eher weniger zeigt, sondern viel mehr auf diese verwiesen wurde. Korpus C zeigt in diesem Zusammenhang vor allem die ziemlich neutrale und nüchterne Betrachtung einer bewaffneten Militanz („they led an armed“, „armed militancy“, „history of religious service“), jedoch keinerlei offensichtliche Bewertungen der Terrorgruppe – weder in positiver, noch in negativer Hinsicht. Anhand dieser beiden Clusteranalysen lässt sich also einerseits anhand von Schlüsselbegriffen sehr einfach der Kontext einordnen, in dem diese Terme stehen, andererseits anhand von n-Grammen gebräuchliche Wortsequenzen filtern, die wiederum Rückschlüsse auf den Sprachgebrauch innerhalb eines Korpus zulassen.

Ein weiterer Indikator, der Rückschlüsse auf sprachliche Eigenschaften eines Korpus zulässt, ist das Auftreten von Kookkurrenzen in Bezug auf bestimmte Terme. Dabei wird auf Basis von Häufigkeiten herausgearbeitet, welche Terme in einem konkreten Wortabstand in Bezug auf einen Suchterm übermäßig gemeinsam auftreten. Auf diese Weise lassen sich Termkombinationen erarbeiten, die eine besonders signifikante kontextuelle Gemeinsamkeit aufweisen. Um auch im Falle dieses Indikators wieder die Vergleichbarkeit zu wahren, wurde erneut auf den bereits verwendeten Term „allah“, aber auch „jihad“ zurückgegriffen. Unter Berücksichtigung der Stoppwortliste (siehe Anhang 4) ergaben sich somit insgesamt sechs Listen mit je 15 Kookkurrenzen (zwei Listen pro Korpus), die auf die beiden Terme bezogen waren (siehe Anhang 9-11). Diese Listen zeigen die 15 stärksten Kookkurrenzen, die besonders häufig gemeinsam mit den beiden Termen auftreten, was in diesem Beispiel vom Suchterm aus mit dem Wortabstand von $k_{\text{links}} = 3$ und $k_{\text{rechts}} = 3$ begrenzt wurde.

Ein Vergleich dieser Listen unterstreicht, was bereits im Rahmen der Clusteranalysen herausgefunden wurde: In Korpus A ist der Anteil an

theologisch, ehrfürchtig und episch konnotierten Wörtern in Verbindung mit dem Term „allah“ deutlich höher, als in Korpus B oder gar Korpus C. Dabei überschneiden sich sogar viele der Kookkurrenzen mit den Sequenzen aus der Clusteranalyse („messenger“, „permission“, „cause“, „grace“). In Korpus B hingegen ist der Anteil an Hashtags („#muslim“, „#muslims“, „#islam“) deutlich höher als noch in der Clusteranalyse, wenngleich auch viele Kookkurrenzen in der Clusteranalyse zu finden sind. Ein wichtiges Merkmal in Korpus B, das sich von der Clusteranalyse jedoch unterscheidet, ist die gruppenspezifische Sprache, die sich durch bestimmte Terme („brother“, „o“ [als direkte Anrede an „allah“]) äußert, eine Konfliktsituation erschafft („victory“ [im Sinne „In-Group“ gegen „out-group“]) und sich mit den gruppendynamischen Kommunikationsprozessen deckt, die bereits oben beschrieben wurden (siehe Kapitel 2.3.3). Korpus C hingegen besticht wieder durch seine vergleichsweise neutralen Kookkurrenzen in Bezug auf den Term „allah“. Wenngleich viele der Top-15-Kookkurrenzen einen theologischen Bezug haben, so sind diese doch eher positiv und harmlos konnotiert. Bezogen auf den Term „jihad“ ergibt sich sogar ein noch klareres Bild, da in Korpus A unter den Top-15-Kookkurrenzen verschiedene Formen des Terms „waging“ mehr als ein Drittel aller Kookkurrenzen ausmachen (siehe Anhang 9). Neben Termen wie „obligation“ und „kuffar“ (als abwertende Bezeichnung für Nicht-Muslime), die die Notwendigkeit der Durchführung des Jihad verdeutlichen sollen, finden sich in den Top-15-Kookkurrenzen auch arabische Begriffe, die ebenfalls islamische Bezüge aufweisen. Dazu gehören „tawhid“ (als islamischer Begriff des Monotheismus), „tawaghit“ (Bezeichnung für die Götter der Feinde Muhammads), „ribat“ (Name der Grenzbefestigungen des islamischen Gebietes zur Durchführung des Jihad) und „shahadah“ (islamisches Glaubensbekenntnis), die allesamt als aggressiv und herabwürdigend konnotiert beschrieben werden können. Anhand dieser Kookkurrenzen wird der aggressive und expansive sprachliche Charakter der

Terrorgruppe besonders deutlich. Die Kookkurrenzen in Korpus B (siehe Anhang 10) fallen vor allem dadurch auf, dass viele Hashtags vorhanden sind, die vom IS im Rahmen seiner Online-Strategie benutzt werden („#tcot“, „#jihadist“, „#gdn“). Interessant ist auch hier wieder der Term „video“, der wieder als Indikator für das häufige Teilen von Propagandavideos gedeutet werden kann. Die Ergebnisse decken sich größtenteils mit denen der Clusteranalyse, weshalb eine Betrachtung des Korpus C sinnvoller ist. Interessant ist die Tatsache, dass in Korpus C (siehe Anhang 11) vor allem negativ konnotierte Terme auftreten, die weder in Korpus A, noch in Korpus B eine Rolle spielen („massacres“, „threatens“). Dies lässt sich als negative Beschreibung der Taten des IS durch neutrale Twitternutzer deuten. Interessant ist auch, dass kontextferne Terme wie „maidan“, „uk“, „trump“ und „law“ auftreten. Die Ursache könnte sein, dass viele Nutzer ein aktiveres Eingreifen der Regierung Trump, beziehungsweise des Vereinten Königreich fordern. Es zeigt sich erneut, dass deutlich größere Gemeinsamkeiten zwischen Korpus A und Korpus B, als zwischen Korpus A und Korpus C existieren.

Eine sehr wirksame Möglichkeit, um die Ergebnisse einer Kookkurrenzanalyse zu visualisieren, ist die Nutzung von Wordclouds. Diese wurden auf Basis der Kookkurrenzen zu jedem Korpus für beide Terme durchgeführt und weisen neben den Top-15-Kookkurrenzen (Anhang 9-11) noch weitere auf (siehe Anhang 12-14). Anhand dieser Abbildungen lässt sich bezogen auf die Schriftgröße der jeweiligen Terme sehr klar die Wichtigkeit dieser in Bezug auf die Ausgangsterme erkennen.

Neben den bereits behandelten Indikatoren spielt auch die sogenannte Keyword List eine besondere Rolle unter den Indikatoren. Diese bestimmt auf Basis statistischer Berechnungen, welche Terme in einem Korpus in Vergleich zu einem Referenzkorpus besonders häufig und besonders selten vorkommen, was einen Hinweis auf Schlüsselwörter liefern kann. Dies ist vor

allem in der vorliegenden Arbeit besonders sinnvoll, da insgesamt drei Korpora existieren und ein Ziel dieser Arbeit darin besteht, den Sprachgebrauch verschiedener TG mit dem in offiziellen IS-Medien (Korpus A) zu vergleichen.

Nachdem also Korpus A als Referenzkorpus geladen wurde, wurde sich in den Einstellungen darauf festgelegt, als statistisches Maß „Chi-Quadrat“ und als kritischen Wert $P = 0,05$ zu wählen. Auf Basis dieser Einstellungen wird dann den Keywords, die besonders häufig und besonders selten vorkommen, ein Wert namens Keyness zugeordnet, der sowohl negative, als auch positive Werte erreichen kann und die jeweilige Korpusgröße berücksichtigt, weshalb nicht in relative Häufigkeiten umgerechnet werden muss. Des Weiteren sollten nur die Top-300-Terme angezeigt werden, da die Ergebnisliste völlig ausreichend ist und auch die nach statistischem Maße negativsten Terme darin enthalten sind. Zunächst wurde Korpus B mit Referenzkorpus A verglichen (siehe Anhang 15) und anschließend Korpus C mit Referenzkorpus A (siehe Anhang 16). Anhand dieser sehr langen Listen ist es möglich, die Breite der genutzten Terme zu erfassen und einen ersten Eindruck zu gewinnen, welche thematischen Schwerpunkte im Vergleich zum Referenzkorpus A in den beiden anderen Korpora vorliegen. Besonders auffällig ist dabei, dass in Korpus B sehr viele höchst negativ konnotierte Terme, die sowohl islamisch-jihadistische Bezüge aufweisen, als auch von allgemeiner Natur sind, häufiger als im Referenzkorpus A genutzt werden. Im Vergleich dazu sind bei den negativen Werten, also den Termen, die seltener als im Referenzkorpus A auftreten, weniger dieser höchst negativ konnotierten Terme dabei, sondern vielmehr solche, die eine islamisch-theologische Konnotation enthalten. Daraus lässt sich schließen, dass der IS im Rahmen seiner Propagandamagazine eine Sprache nutzt, die deutlich zurückhaltender ist als die, die im Rahmen der sozialen Medien verwendet wird. Eine Ursache könnte darin liegen, dass in den Magazinen ein professionelles Bild der

Gruppe mit theologischem Anspruch erschaffen werden soll, wie dies auch bereits in anderen Arbeiten erörtert wurde (siehe Kapitel 2.3.1). Der Vergleich zwischen Korpus C und Referenzkorpus A hingegen zeichnet ein anderes Bild, wie sich anhand der Breite der genutzten Terme zeigt. Dort sind neben einigen größtenteils neutralen islamisch-theologischen Termen vor allem solche zu finden, die die aktuelle Nachrichtenlage abbilden und keinen theologischen Kontext aufweisen. Auffällig ist ebenso das im Vergleich zu Referenzkorpus A häufige Auftreten von positiven Adjektiven, während unter den Termen, die deutlich seltener vorkommen, einige islamisch-jihadistische Terme vorzufinden sind. Dies zeigt sehr deutlich, dass im Rahmen von Korpus C, also der Gruppe der allgemeinen Twitternutzer, vor allem eher distanziert, negativ und verknüpft zu anderen Themen über die Gruppe geschrieben wird.

Am deutlichsten werden die Unterschiede, wenn konkrete Terme betrachtet werden, wie sich am Beispiel von „quran“ besonders deutlich zeigt. Während in Korpus B dieser Term nur marginal häufiger (+154.25) vorkommt als im Referenzkorpus A, so spielt dieser Term in Korpus C eine deutlich wichtigere Rolle (+2435.01) als im Referenzkorpus. Noch deutlicher wird dieser Umstand bei der Betrachtung des Terms „jihad“, der in Korpus B deutlich häufiger (+5144.63) genutzt wird in Bezug auf Korpus A, als es bei Korpus C (-68,63) der Fall ist, wo dieser Term sogar seltener als im Referenzkorpus A genutzt wird. Diese Werte weisen darauf hin, dass die TG in Korpus C einen viel engeren Bezug zum Koran aufweist, als die TG in Korpus B – diese hingegen jedoch eine deutlichere sprachliche Affinität zu radikalen Termen wie „jihad“ zeigt, was angesichts der Gruppenstruktur jedoch auch nicht unerwartet ist.

Diese Vielzahl von korpuslinguistischen Untersuchungen zeigt vor allem, dass zwischen Korpus B und dem Referenzkorpus A eine deutlich größere sprachliche Nähe existiert, als zwischen Korpus C und A. Diese Indikatoren,

die unabhängig voneinander anhand verschiedener Methoden sprachliche Eigenschaften skalieren, machen diese Gemeinsamkeiten und Unterschiede nicht nur messbar, sondern auch vergleichbar. Auf diese Weise konnte gezeigt werden, in welchem Verhältnis die Korpora zueinander stehen und welche Charakteristika diese aufweisen.

3.5 Weitergehende Analysen

Neben den bereits durchgeführten korpuslinguistischen Analysen gibt es auch noch weitere Verfahren, die Rückschlüsse auf die Beschaffenheit und Charakteristik von Texten geben. Von großer Relevanz ist dabei das Verhältnis des Auftretens von besonders aussagekräftigen Termen zwischen den jeweiligen Korpora. Während in den korpuslinguistischen Analysen bisher verschiedene Indikatoren betrachtet wurden, die deskriptive Ergebnisse geliefert haben, soll nun im folgenden Schritt die mathematische Verifizierung dieser mittels Signifikanztest erfolgen. Konkret steht dabei die Frage im Raum, ob mit ausreichend hoher Wahrscheinlichkeit gesagt werden kann, dass die unterschiedlichen Frequenzen von besonders aussagekräftigen Termen nicht rein zufällig entstanden sind, sondern vielmehr bewusst verwendet wurden. Mathematisch wird dieser Signifikanztest in Form eines Chi-Quadrat-Tests durchgeführt, da neben dem Chi-Quadrat-Test auch das Signifikanzniveau der einzelnen Testreihen miteinander verglichen werden kann.

Zunächst müssen dafür Kontingenztabellen für jeden Term aufgestellt werden, der sowohl im Referenzkorpus A, als auch im jeweiligen Vergleichskorpus enthalten sein muss. Aus diesem Grund werden die Signifikanztests anhand der Terme durchgeführt, die in absoluten Zahlen unten zu finden sind (siehe Anhang 15 und 16). Als Testparameter dienen die fünf Terme „allah“, „state“, „prophet“, „jihad“ und „islam“, die allesamt in jedem der drei Korpora zu finden sind. Der Aufbau der Kontingenztabellen

gleicht immer einem Muster: Zunächst werden die beobachteten Frequenzen (O) eines Terms in jedem Korpus extrahiert und diese anschließend mit allen anderen Wörtern eines Korpus zusammen aufgelistet, wie dies beispielhaft an „allah“ in Korpus B gezeigt wird (Tab. 4).

	Korpus A	Vergleichskorpus B	Total
Frequenz Term „allah“	7674	490	8164
Alle anderen Wörter	700878	185879	886757
Total	708552	186369	894921

Tabelle 4: *Kontingenztafel beobachteter Werte des Terms „allah“* (Eigene Darstellung)

Anschließend kann die Nullhypothese (H_0) definiert werden, die im folgenden Beispiel lautet: *Die Korrelation zwischen der Termfrequenz und dem Korpus ist zufällig zustande gekommen.* Die Gegenhypothese (H) hingegen lautet: *Die Korrelation zwischen der Termfrequenz und dem Korpus ist nicht zufällig zustande gekommen.* Um mittels Signifikanztest herauszufinden, ob H_0 verworfen werden kann, müssen im nächsten Schritt anhand der Randhäufigkeiten die erwarteten Werte (E) errechnet werden, was mittels Dreisatz schnell nach folgendem Prinzip erledigt werden kann (Tab. 5).

	Korpus A	Vergleichskorpus B	Total
Frequenz Term „allah“	$\frac{708552 * 8164}{894921}$ = 6464	$\frac{186369 * 8164}{894921}$ = 1700	8164
Alle anderen Wörter	$\frac{708552 * 886757}{894921}$ = 702088	$\frac{186369 * 886757}{894921}$ = 184669	886757
Total	708552	186369	894921

Tabelle 5: *Kontingenztafel erwarteter Werte des Terms „allah“* (Eigene Darstellung)

Vor der Durchführung des Tests ist es noch wichtig, den Freiheitsgrad df zu berechnen, der sich aus $(\text{Reihenzahl} - 1) * (\text{Spaltenzahl} - 1)$, also in den vorliegenden Fällen $df = 1$ ergibt. Anschließend wird die Chi-Quadrat-Prüfgröße χ^2 bestimmt, die sich wie folgt errechnet:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Daraus folgt:

$$\chi^2 = \left(\frac{(7674 - 6464)^2}{6464} \right) + \left(\frac{(490 - 1700)^2}{1700} \right) + \left(\frac{(700878 - 702088)^2}{702088} \right) + \left(\frac{(185879 - 184669)^2}{184669} \right)$$

$$\chi^2 = 226,5 + 861,24 + 2,09 + 7,93$$

$$\chi^2 = 1097,76$$

Anhand einer Tabelle (siehe Anhang 17) mit kritischen Werten für χ^2 lässt sich anschließend überprüfen, ob der errechnete Wert signifikant ist. Da

aufgrund des Freiheitsgrades $df = 1$ für die vorliegenden Rechnungen nur die erste Zeile der Tabelle von Bedeutung ist, lässt sich schnell erkennen, dass der Wert $\chi^2 = 1097,76$ sogar so hoch ist, dass ein Signifikanzniveau von $P = 0,001$ erreicht wird, obwohl im Regelfall bereits ein Signifikanzniveau von $P = 0,05$ ausreicht, um H_0 zu verwerfen. Im Umkehrschluss bedeutet dies, dass mit 99,9% Wahrscheinlichkeit davon ausgegangen werden kann, dass die Frequenzunterschiede nicht zufällig zustande gekommen sind. In Bezug auf den Term „allah“ lässt sich also mit 99,9% Wahrscheinlichkeit sagen, dass in Korpus B bewusst die Nutzung des Terms reduziert wurde, während sie in Korpus A leicht häufiger verzeichnet wurde.

Um im Folgenden den mathematischen Prozess zu vereinfachen, wurden die einzelnen Schritte im Rahmen eines Tabellenkalkulationsprogrammes durchgeführt, was zu folgenden Ergebnissen führte (Tab. 6). Die Daten dazu befinden sich unten (siehe Anhang 18).

Terme	χ^2 in Korpus B	χ^2 in Korpus C
„allah“	1097,76	207,54
„state“	31,06	4,73
„prophet“	111,07	589,9
„jihad“	4559,26	453,62
„islam“	1681,51	142,48

Tabelle 6: *Chi-Quadrat-Prüfwerte der Terme in Bezug auf Referenzkorpus A* (Eigene Darstellung)

Für Korpus C lässt sich bei der Nutzung des Terms „allah“ ebenfalls feststellen, dass ein Signifikanzniveau von $P = 0,001$ erreicht wurde. Jedoch findet die Nutzung dieses Terms in Korpus C häufiger statt als in Korpus A, weshalb davon ausgegangen werden kann, dass dieser in der Gruppe der allgemeinen Twitternutzer eine gewichtigere Rolle spielt als im IS-

Referenzkorpus. Aufschlussreich ist auch die Betrachtung des Terms „state“, der bei Korpus B ein Signifikanzniveau von $P = 0,001$ erreicht, bei Korpus C jedoch lediglich von $P = 0,05$. In Korpus B wird dieser Term dabei seltener verwendet als im Referenzkorpus, was ein Indiz dafür ist, dass die staatsbildende Propaganda in den Magazinen von größerer Bedeutung ist, als es bei den Akteuren auf Twitter der Fall ist. Aufgrund der relativ vergleichbar häufigen relativen Nutzung des Terms in Korpus A und Korpus C ist das Signifikanzniveau aus diesem Grunde auch relativ niedrig, weshalb Interpretationen sich ebenfalls deutlich schwieriger anhand dieses Indikators gestalten. Anders ist die Situation jedoch beim Term „prophet“, der im Referenzkorpus A relativ gesehen so viel häufiger vorkommt als in Korpus B und Korpus C, dass sich dies auch merklich im Signifikanzniveau niederschlägt. Bei beiden Korpora ergibt sich in Bezug auf Korpus A ein Signifikanzniveau von $P = 0,001$, was nochmal anschaulich zeigt, wie theologisch aufgeladen die Magazine sind. Die deutlichsten Interpretationen lässt jedoch der Term „jihad“ zu, wie sich auch anhand der Signifikanzniveaus zeigt, die bei Korpus B und Korpus C jeweils bei $P = 0,001$ liegen. Interessant ist jedoch, dass die Nutzung des Terms in Korpus B deutlich häufiger als im Referenzkorpus stattfindet, was ausdrücklich die bereits gesammelten Ergebnisse bestätigt, die oben bereits beschrieben wurden (siehe Kapitel 3.4). Der Fokus bei den IS-Aktivisten auf Twitter liegt in der Propagierung des Jihad, während dieser in den Magazinen eine eher unterschwellige Position einnimmt. In Korpus C hingegen spielt dieser Term in Bezug auf den Referenzkorpus eine geringe Rolle, was jedoch umso deutlicher wird, wenn man bedenkt, dass selbst im Referenzkorpus dieser Term relativ unwichtig ist. Das Signifikanzniveau dürfte also nochmals deutlich höher sein, wenn man den Chi-Quadrat-Test zwischen Korpus B und Korpus C durchführen würde. Als letzter Term wurde „islam“ betrachtet, dessen statistische Analyse ebenfalls wieder eindeutige Signifikanzniveaus

von $P = 0,001$ in beiden Korpora zeigte. Während dieser in Korpus B eine geringere Rolle als im Referenzkorpus spielt, ist seine Bedeutung in Korpus C deutlich größer, was wiederum zeigt, dass die IS-Aktivisten auf Twitter wenig Wert auf theologischen Inhalt legen. In Korpus C hingegen ist der Term von deutlich größerer Wichtigkeit, als im Referenzkorpus, was wiederum zeigt, dass der IS – trotz seiner vermeintlichen Versuche, sich einen theologischen Anstrich zu geben – im Vergleich zu allgemeinen Twitternutzern weniger Wert auf diesen Schlüsselbegriff legt.

Wie bereits bei den Indikatoren zuvor zeigt sich auch im Rahmen einer statistischen Betrachtung von Schlüsselwörtern, dass diese Analysen viele Informationen liefern können, jedoch immer auch vergleichende Daten erhoben werden müssen, da die Signifikanzniveaus ohne Vergleichsindikatoren keine Aussagekraft besitzen. Um diese Ergebnisse jedoch auch nochmal aus einer eher semantischen Perspektive heraus zu betrachten, soll im Folgenden als letzte Ergänzung eine Sentimentanalyse durchgeführt werden, die die allgemeine Haltung der 1000 meistfrequentierten Terme der einzelnen Korpora zusammenfassen soll. Diese nimmt dabei jedoch nicht die Rolle einer alternativen, sondern vielmehr einer ergänzenden Analyse ein. Dabei werden die Terme daraufhin untersucht und gewichtet (von -5 bis +5), wie diese konnotiert sind, um die Haltung dieser zu skalieren und vergleichbar zu machen.

Wie sich anhand der graphischen Darstellung der Sentimentscores der Top-1000 Terme in den einzelnen Korpora zeigt (siehe Anhang 19-21), existieren auch Unterschiede in der Konnotation der genutzten Terme. Besonders auffallend ist, dass in Korpus C, im Gegensatz zu Korpus A und Korpus B, ein deutlich häufigeres Auftreten von Termen mit positiven Sentimentscores vorzufinden ist. Dieser Korpus ist des Weiteren auch der einzige, der bei den positiven Scores den Wert von +4 erreicht, während Korpus A und B lediglich ein Maximum von +3 erreichen. Im negativen Wertebereich ist

Korpus B derjenige, der das Werteminimum von -5 als einziger Korpus erreicht, während Korpus A und C lediglich bis -4 reichen. Gerade durch die Breite der negativen Werte lässt sich feststellen, dass Korpus B derjenige ist, der insgesamt den negativsten Sentimentscore erreicht. Ein Vergleich der jeweiligen Sentimentscores im Rahmen der Graphiken zeigt demnach erneut, dass zwischen Korpus A und Korpus B eine deutlich größere Nähe existiert, als zwischen Korpus A und Korpus C. Auch dieser Indikator zeigt unabhängig von den anderen, dass der Sprachgebrauch in den Korpora A und B miteinander enger verbunden ist.

4 Diskussion und Fazit

Im letzten Teil dieser Arbeit wird nochmals Bezug auf die bisher gewonnen Ergebnisse genommen, wobei diese besonders im Spiegel der Fragestellungen betrachtet werden, die zu Beginn definiert wurden. Dabei ging es konkret um die Frage, ob sich Personen mit Zuneigung zu einer bestimmten Ideologie anhand ihres Sprachgebrauchs identifizieren lassen und ob es Indikatoren gibt, die dafür besonders ergiebig sind. Außerdem standen auch die Unterschiede und Gemeinsamkeiten zwischen den jeweiligen Korpora und damit zwischen den unterschiedlichen TG im Fokus.

Die bereits gesammelten Ergebnisse sollen dazu nochmals kurz zusammengefasst und in einen gemeinsamen Kontext gebracht werden, wobei die korpuslinguistischen Analysen in einem Block gesondert betrachtet werden, während die statistischen Betrachtungen und die SA in einem zweiten Block behandelt werden. Im daran anschließenden Schritt sollen dann auch Möglichkeiten genannt werden, wie eine solche Arbeit weitergeführt, beziehungsweise optimiert werden könnte und welche weitergehenden Erkenntnisse bei der Durchführung der Arbeit aufgetreten sind. Im letzten Teil werden dann abschließend die wichtigsten Ergebnisse nochmals zusammengefasst präsentiert.

4.1 Korpuslinguistische Sprachanalyse

Wie der Vergleich der relativen Häufigkeiten von mehreren Termen aus Tabelle 3 zeigt, existieren einige Gemeinsamkeiten zwischen den jeweiligen Korpora. Auffallend ist dabei besonders, dass vermeintlich neutrale Terme wie „allah“, „state“, „islamic“, „people“ oder „muslims“ eine besonders enge Gemeinsamkeit zwischen Referenzkorpus A und dem Korpus C aufzeigen. Werden jedoch noch weitere spezifische Terme hinzugezogen, die in den Korpora häufig verwendet werden, so werden die Unterschiede in der Verwendung der jeweiligen Terminologie besonders deutlich. Korpus A fällt

vor allem durch die Verwendung theologischer Terme auf, was in Verbindung mit dem Anspruch, die einzig wahre Deutung des Korans zu praktizieren, plausibel scheint. Des Weiteren spiegeln sich auf diese Weise auch die Propagandafaktoren wieder, die in 2.3.1 bereits behandelt wurden, was auch für den Anspruch auf das Monopol des Regimes zur Gewaltausübung in göttlichem Einsatz gilt. Korpus B hingegen zeigt einen deutlich aggressiveren Sprachgebrauch, der sich vor allem durch die Nutzung von Termen wie „jihad“, „sharia“ und verschiedenen Bezeichnungen für die eigene Gruppe und das eigenen Gebiet zeigt. Dieser Umstand verdeutlicht die Absicht der Rekrutierung durch den IS, was ebenso auf Aspekte wie Staatsbildung und die Nutzung von Twitter als Megaphon für Propagandaaktivitäten zutrifft und in 2.3.2 behandelt wurde. In Korpus C, der ebenso wie die beiden anderen Korpora aufgrund der thematischen Nähe theologische Aspekte und den IS behandelt, fällt hingegen besonders die vielfache Nutzung von positiv konnotierten Termen auf, die in den beiden anderen Korpora eine deutliche geringere, wenn nicht sogar gar keine Rolle spielen.

Anhand von Konkordanz-Plots lässt sich sehr anschaulich die Verwendung von Termen untersuchen, so wie es in dieser Arbeit anhand von „allah“ und „state“ in den Propagandamagazinen geschehen ist. Dabei fällt auf, dass sich die hohe Konzentration von „allah“ vor allem zur militärischen Hochzeit und zu Zeiten größter militärischer Bedrängnis in den Propagandamagazinen widerspiegelt. Das lässt den Schluss zu, dass zur Hochzeit Allah als religiöses Motiv, zur Zeit der Bedrängung hingegen als propagandistische Durchhalteparole genutzt werden sollte. Die Betrachtung des Terms „state“ unterstützt diese Sichtweise, da dieser lediglich zu Zeiten von militärischen Siegen des IS seine Verwendung fand, während er später, als das selbsternannte Kalifat schon fast verloren war, kaum noch Berücksichtigung in den Magazinen fand (siehe auch Abb. 5).

Auch die Clusteranalysen unterstützen die Ergebnisse, die schon im Rahmen der Häufigkeitsanalysen gewonnen wurden. Auffallend ist dabei, dass eine große Ähnlichkeit in der Verwendung von Wortsequenzen zwischen den Korpora A und B existiert, die die Mächtigkeit des Gottes verdeutlichen sollen, was sich auch durch ehrfürchtige Ansprachen in der dritten Person ausdrückt. In Korpus C hingegen wenden sich die Nutzer mit persönlichen Wünschen an ihren Gott und nutzen dafür meist die direkte Anrede. Dieser Aspekt verdeutlicht den Unterschied zwischen gemäßigtem Glauben an einen Gott, wie er sich in Korpus C ausdrückt, und dem religiösen Fundamentalismus, für den der IS steht und welcher sich in den Korpora A und B zeigt (siehe Anhang 7). Die termunabhängigen Cluster (siehe Anhang 8) bestätigen dann nochmals die thematischen Schwerpunkte, die sich bei Korpus A auf militärisch-theologische und häufig in Arabisch verfasste Cluster beschränken, während in Korpus B unverhohlen direkt der Jihad und die Radikalisierung thematisiert werden, während in Korpus C eher beschreibende Wortcluster zu finden sind.

Bei der Betrachtung der Kookkurrenzen fällt vor allem auf, dass Korpus A in Bezug auf den Term „allah“ neben theologischen auch arabische Wörter beinhaltet, die der islamischen Glaubenslehre entstammen. Dies macht erneut deutlich, dass sich vonseiten des IS der Anspruch der wahren Deutung des Islam sprachlich niederschlägt. Dabei scheint, was die Kookkurrenzen vom Term „jihad“ zeigen, die Durchführung dessen aus deren Sicht ein probates Mittel zu sein (siehe Anhang 9). Ein ähnliches Muster zeigt sich auch in Korpus B (siehe Anhang 10), der ähnliche Kookkurrenzen hat, wenngleich aufgrund der Korpuseigenschaften das Auftreten von Hashtags häufiger vorkommt. Korpus C (siehe Anhang 11) hingegen beinhaltet Kookkurrenzen, die positive Eigenschaften in Bezug auf „allah“ aufweisen, den Term „jihad“ jedoch vor allem mit negativen Bezügen kennzeichnen. Diese Kookkurrenzen verdeutlichen also ebenso sehr eindeutig die Unterschiede, die zwischen den

radikalen Sprachmustern in den Korpora A und B und den neutralen Ausdrücken in Korpus C existieren. Diese Unterschiede sind auch graphisch erfassbar, wie die Kookkurrenz-Wordclouds zeigen (siehe Anhang 12-14).

Die Keyword Lists, die sich vor allem dadurch auszeichnen, Testkorpora auf einen Referenzkorpus anzuwenden und anhand dessen Terme zu analysieren, die sowohl besonders häufig, als auch seltener im Vergleich zum Referenzkorpus vorkommen, bestätigen ebenso diese Ergebnisse. In Korpus B ist demnach auffällig, dass dieser viele negativ konnotierte Terme beinhaltet, die Bezug auf Gewalt nehmen, während vermeintlich neutralere Terme eher seltener als im Referenzkorpus A vorkommen. Dies lässt den Schluss zu, dass die Anhänger der IS-Ideologie, beziehungsweise deren Akteure in sozialen Netzwerken deutlich freier und direkter kommunizieren, als es in den offiziellen Propagandamagazinen der Fall ist. Im Endeffekt unterstreicht dies die Absicht, offiziell seriös wirken, aber in sozialen Netzwerken aktiv und aggressiv Rekrutierung betreiben zu wollen, wie dies auch in 2.3.2 beschrieben wurde. Korpus C hingegen ist in seiner Sprache deutlich zurückhaltender und positiver konnotiert als Referenzkorpus A. Besonders extreme Terme wie „jihad“ werden sogar deutlich weniger genutzt, was wiederum die Unterschiede zwischen den allgemeinen Twitternutzern und den Fundamentalisten aufzeigt (siehe Anhang 15 und 16).

Im Rahmen der korpuslinguistischen Analysen wurde deutlich, dass über mehrere Indikatoren unabhängig voneinander größere Ähnlichkeiten zwischen den Korpora A und B bestehen, während Korpus C in jeder Hinsicht eine eher zurückhaltende, teils höchst konträre Nutzung der Sprache repräsentiert. Anhand dieser Indikatoren lässt sich also durchaus sagen, dass große Unterschiede zwischen allgemeinen Twitternutzern und dem IS, beziehungsweise seinen Sympathisanten existieren. Des Weiteren lassen sich sogar in Bezug auf die Verwendung bestimmter Wortcluster enge Gemeinsamkeiten zwischen Korpus A und Korpus B nachweisen, wobei vor

allem die Clusteranalysen, die Kookkurrenzen und die Keyword Lists eindeutige Ergebnisse geliefert haben. Wenngleich eine gewisse thematische Nähe in Bezug auf theologische Terme besteht, lassen sich dennoch anhand dieser Indikatoren Gebrauchsmuster feststellen, die sich eindeutig jihadistischer Ideologie zuordnen lassen und die sich stark vom Sprachgebrauch in Korpus C – also dem der allgemeinen Twitternutzer – unterscheiden lassen. Somit können diese Indikatoren sinnvoll zur Identifikation von IS-Sympathisanten genutzt werden, was auch auf die Detektion von LW zutrifft.

4.2 Weitergehende Analysen

Auch in Bezug auf die statistischen Betrachtungen der Korpora lassen sich einige Besonderheiten feststellen. Dafür wurde das Auftreten von fünf verschiedenen und besonders aussagekräftigen Termen in Korpus B und Korpus C in Bezug auf Referenzkorpus A statistisch untersucht. Dabei wurde bereits im Rahmen der Keyword Lists herausgearbeitet, dass sich das Auftreten dieser Terme in den Korpora B und C stark von dem in Referenzkorpus A unterscheidet. Wenngleich es höchst unwahrscheinlich ist, dass dies in allen zehn untersuchten Fällen zufällig geschehen ist, wurde trotzdem ein Chi-Quadrat-Test durchgeführt, um mithilfe mathematischer Signifikanz und daraus folgend mit ausreichend hoher Wahrscheinlichkeit davon ausgehen zu können, dass diese Auftretungshäufigkeiten nicht zufällig entstanden sind.

Bei der Durchführung der Chi-Quadrat-Tests zeigte sich dabei, dass in neun von zehn Fällen ein Signifikanzniveau von $P = 0,001$ erreicht wurde, also mit 99,9% Wahrscheinlichkeit davon ausgegangen werden kann, dass die Frequenzunterschiede nicht zufällig, sondern absichtlich zustande gekommen sind. Lediglich in einem Fall („state“ in Korpus C) belief sich das Signifikanzniveau auf $P = 0,05$, was jedoch immer noch ein aussagekräftiges

Ergebnis ist, da mit 95% Wahrscheinlichkeit von einem nicht zufälligen Zustandekommen ausgegangen werden kann. Diese statistischen Signifikanztests in Form von Chi-Quadrat-Tests haben mathematisch bewiesen, dass ein häufigeres, beziehungsweise selteneres Auftreten von den betrachteten Termen bewusst provoziert wurde.

Wie auch bereits die Indikatoren im vorherigen Kapitel zeigten, wurde beispielsweise der Term „allah“ in Korpus B deutlich seltener verwendet als im Referenzkorpus A, was mutmaßlich damit zusammenhängt, dass dem IS zwar offiziell in den Magazinen eine religiöse Erscheinung zugeschrieben werden soll, diese Akteure sich aber zeitgleich in sozialen Netzwerken sprachlich weniger religiös, als vielmehr fundamentalistisch ausdrücken, um etwa Rekrutierungen durchzuführen oder Hassbotschaften zu verbreiten, während hingegen Korpus C, der allgemeine Twitternutzer abbildet, deutlich häufiger eine Nutzung des Terms „allah“ beinhaltet, was jedoch auf die religiösen Bittstellungen und Gebete der Nutzer zurückzuführen ist, die mit jihadistischem Terror wenig gemeinsam haben.

Ein Blick auf den Term „state“ offenbart ebenso, dass die Nutzung in Korpus B eine geringere Rolle als im Referenzkorpus A spielt, da staatsbildende Propaganda eher in den Magazinen als in den Profilen der Aktivisten in sozialen Medien zu finden ist. Der Term „prophet“ kommt sowohl in Korpus B, als auch in Korpus C so viel seltener als im Referenzkorpus A vor, dass dadurch der vermeintlich theologische Anstrich der Magazine sehr gut verdeutlicht wird.

Ebenso bezeichnend ist die Nutzung des Terms „jihad“, die eine außergewöhnlich große Rolle in Korpus B einnimmt, was angesichts der Hassbotschaften (siehe 2.3.2), Gruppendynamiken (siehe 2.3.3) und Ziele (siehe 2.3.1) der Terrorgruppe wenig überraschend ist. Verstärkt wird dieser Effekt dadurch, dass die Nutzung des Terms „jihad“ in Korpus C noch seltener vorkommt als in Referenzkorpus A, obwohl die Nutzung dort bereits

deutlich seltener als in Korpus B stattfindet. Dies verdeutlicht sehr anschaulich den aggressiven Sprachcharakter, der in Korpus B vorzufinden ist und pure Gewalt predigt. Ein letztes Gegenbeispiel hingegen ist die Nutzung des Terms „islam“, die in Korpus B deutlich seltener, in Korpus C hingegen häufiger stattfindet. Damit wird wieder gezeigt, dass neutrale religiöse Motive in Korpus C deutlich wichtiger sind als in Korpus B.

Die statistischen Signifikanztests beweisen demnach nicht nur, dass die Frequenzen der aufgetretenen Terme nicht rein zufällig zustande gekommen sind, sondern stehen dabei auch in direktem Kontext zu den aus der korpuslinguistischen Analyse gewonnenen Ergebnissen der einzelnen Indikatoren. Eine Kombination dieser Indikatoren mit mathematischen Signifikanztests ermöglicht dabei, genaue Annahmen auf Basis mathematischer Signifikanz unter vergleichsweise geringem Aufwand treffen zu können, beziehungsweise gewonnene Ergebnisse verifizieren zu können.

Auch durch die abschließende Betrachtung der Top-1000 Terme der jeweiligen Korpora und deren Anwendung im Rahmen einer SA, die abgesehen von der Frequenzliste absolut unabhängig von den korpuslinguistischen oder statistischen Analysen ist, wurde der Gesamteindruck bestätigt (siehe Anhang 19-21).

Dabei ließ sich feststellen, dass Korpus B mit Abstand am häufigsten in den negativen Wertebereich reicht und dabei gleichzeitig auch das Werteminimum als einziger Korpus erreicht, was folglich für seinen besonders negativ konnotierten Sprachgebrauch spricht. Korpus C hingegen ist deutlich positiver angelegt, was auch durch das Erreichen des Wertes +4 als einziger der drei Korpora ausgedrückt wird. Durch den graphischen Plot der Sentimentscores lässt sich außerdem auch visuell erfassen, dass eine deutlich größere Nähe zwischen den Korpora A und B besteht, während C eher als Außenseiter zu betrachten ist.

4.3 Weitergehende Erkenntnisse und zukünftige Möglichkeiten

Im Rahmen dieser Arbeit ergibt sich eine Vielzahl von Erkenntnissen, die infolge der korpuslinguistischen Analysen aufgetreten sind und die im Folgenden für zukünftige Arbeiten vorgestellt werden sollen. Außerdem werden auch weiterführende Ansätze betrachtet, die Arbeiten dieser Art verbessern, beziehungsweise Ergebnisse optimieren könnten.

Besonders wertvoll ist die Durchführung ähnlicher Ansätze, da diese nicht nur – wie in der vorliegenden Arbeit – auf islamistische Terrorgruppen anwendbar ist, sondern theoretisch auf jede Art von Online-Extremismus. Grundvoraussetzung ist dabei stets ein geeigneter Referenzkorpus, von dem aus diese Analysen durchgeführt werden können. Dabei wird sich der Umstand zunutze gemacht, dass die meisten Formen von Online-Extremismus dieselben Charakteristiken aufweisen, die auch im hier angewandten Gruppenkontext aufgetreten sind. So werden mutmaßlich auch andere extremistische Gruppen sowohl nach der Sperrung ihres Profils wieder neue Profile erschaffen, um ihre textliche Propaganda zu verbreiten, als auch wieder eine neue Verknüpfung mit alten Profilen von Mitstreitern anstreben (Klausen, Marks, & Zaman, 2016, S. 36).

Eine weitere, vielleicht gar wichtigste Erkenntnis im Rahmen dieser Arbeit ist der Eindruck, dass das Analysepotential im Rahmen der Korpuslinguistik gefühlt unendlich groß zu sein scheint. Zu jedem einzelnen der behandelten Indikatoren könnte bei gleichem vorhandenem Korpusmaterial eine eigene Arbeit geschrieben werden, zumal diese Indikatoren auch noch auf mehrfache Weise kombiniert werden könnten. Deutlich wird dies besonders, wenn sich bewusst gemacht wird, dass die meisten der hier durchgeführten Analysen nur auf wenigen Termen beruhen, obwohl zehntausende verschiedene ebendieser Terme existieren. Es darf also angenommen werden, dass die indikatorspezifische Betrachtung einer größeren Stichprobe auch zu deutlich genaueren Ergebnissen gelangt. Jedoch zeigt sich auch, dass eine

Kombination mehrerer Indikatoren mit geringer Stichprobe ebenfalls zu stimmigen und plausiblen Ergebnissen führen kann und zur Gewinnung von wichtigen Erkenntnissen ausreichend sein kann. Besondere Beachtung sollte dabei der Umstand erfahren, dass diese Indikatoren unterschiedliche Verfahrensweisen repräsentieren – die Ergebnisse also unabhängig voneinander entstehen, aber trotzdem stimmig sind. Auch die Kombination dieser Indikatoren mit Sentimentanalysen und statistischen Signifikanztests zeigt, dass eine Kombination mehrerer Verfahren die Ergebnisse multiperspektivisch stärken und verifizieren kann.

Jedoch zeigt sich auch erhebliches Verbesserungspotential in mehrerlei Hinsicht. So bestehen die Korpora in dieser Arbeit ausschließlich aus englischen Texten, die hin und wieder mit arabischen Transliterationen durchzogen sind. Obwohl dies aus technischer Sicht eher zu einer Verbesserung der Ergebnisse führt, da englischsprachige Systeme besser elaboriert sind, kann angenommen werden, dass das Korpusmaterial von arabischsprachigen Personen noch signifikantere Ergebnisse liefern würde. Gestützt wird diese Vermutung dadurch, dass im Rahmen der Datenbereinigung leider viele radikale arabischsprachige Textabschnitte entfernt werden mussten, um möglichst homogene Korpora für die Analysen zu nutzen. Eine multilinguale korpuslinguistische Betrachtung könnte – gerade in Bezug auf diesen Themenkomplex – sicher eine ganz neue analytische Dimension eröffnen und Interessante Vergleiche in Bezug auf unterschiedliche Sprachgruppen liefern. Des Weiteren ist die größtenteils englischsprachige Behandlung dieses Themenkomplexes allgemein eine Problematik, die sich im Rahmen vieler Arbeiten wiederfindet und dazu führt, dass die Stichproben beispielsweise nicht als repräsentativ für arabische Personen angesehen werden dürfen, da viel Material nicht genutzt werden kann (Huey, Inch, & Peladeau, 2019, S. 460).

Ein weiterer Punkt, der mithilfe korpuslinguistischer Analysen weiter ausgeführt werden könnte, ist die rollenspezifische Analyse, die die Detektion von konkreten Rollentypen ermöglicht. Obwohl das Online-Verhalten dieser Akteure bereits sehr genau klassifiziert ist, so fand die Anwendung dieser Muster beispielsweise in Detektionssystemen bisher wenig Beachtung, obwohl in dieser Hinsicht korpuslinguistische Analysen sicherlich einen großen Beitrag leisten könnten. Gleiches gilt auch für andere Informationen, die im Rahmen von Netzwerkanalysen gewonnen werden könnten. Auch die Nutzung von Sentimentanalysen könnte noch genauer auf diesen Themenkomplex zugeschnitten werden, wenngleich sich einige grundsätzliche Schwierigkeiten mit SA ergeben. So können beispielsweise keine visuellen Inhalte, die sich in Magazinen oder auf Twitter finden lassen, in die Auswertung mit aufgenommen werden (Macnair, & Frank, 2018a, S. 118). Bisher fand eine SA in diesem Themenkomplex meist Anwendung, um Magazine genauer zu untersuchen, obwohl beispielsweise eine zeitkritische Betrachtung der Propagandamagazine in Kombination mit dem Sentimentscore von IS-affinen Gruppen deutlich interessantere Ergebnisse liefern könnte, da dieser Ansatz nicht nur als weiterer Detektionsindikator, sondern auch als deskriptiver Faktor einer Analyse über die Zeit genutzt werden könnte.

Allgemein betrachtet bieten diese Formen der korpuslinguistischen Analysen ein schier unerschöpfliches Werkzeug, das auf unterschiedlichste Weise genutzt werden kann. So entstand im Rahmen dieser Arbeit mehrmals die Situation, dass eine Entscheidung getroffen werden musste, wie groß der Stichprobenumfang werden soll und welche weiteren Analysen durchgeführt werden sollen, um den Umfang dieser Arbeit nicht allzu groß werden zu lassen. In besonderem Maße traf dies auf die korpuslinguistischen Analysen zu, bei denen aufgrund des begrenzten Umfangs der Arbeit auf weitere Analysen verzichtet werden musste. Obwohl die meisten der hier

durchgeführten Analysen von kleinem Stichprobenumfang sind, zeigen sich trotzdem starke Ergebnisse, die zur Identifizierung von ZP dienen können, da diese mithilfe mehrerer Indikatoren an ihrem Sprachgebrauch messbar und damit identifizierbar sind.

5 Zusammenfassung

Allgemein zeigt sich im Rahmen dieser Arbeit, dass die Bedrohung durch terroristische Gruppen jeglicher Art, von denen der sogenannte „Islamische Staat“ exemplarisch genannt werden kann, nach wie vor eine große Rolle spielt. Aufgrund der großen Wirkung, die diese Terrorgruppe im Laufe der letzten Jahre entwickelt hat, besteht durchaus die reelle Gefahr einer Vorbildfunktion für andere Gruppen. Diese wird mutmaßlich – nicht zuletzt aufgrund der nachweisbaren Erfolge bei der Rekrutierung von Neumitgliedern – auch dazu führen, dass das Internet sich zu einem immer wichtiger werdenden Handlungsraum für terroristische Gruppen entwickeln wird. Dies zeigen zudem auch bereits die letzten Jahre, in denen sich terroristische Gruppen mehr und mehr in sozialen Netzwerken und anderen Plattformen ausbreiteten. Demzufolge wächst auch das Bedürfnis nach technischen Möglichkeiten des Online-Monitorings, um sinnvolle Strategien für CT erstellen zu können.

Neben klassischen Varianten des OM haben sich korpuslinguistische Analysen als ein besonders wirkungsmächtiges Werkzeug herausgestellt, bei denen sich jedoch neben dem Bedarf nach ausreichendem Korpusmaterial auch technische und linguistische Problemstellungen ergeben. Aufgrund der Tatsache, dass die Terrorgruppe IS neben den eigens produzierten Propagandamagazinen auch weite Aktivitäten auf sozialen Netzwerken wie Twitter zeigt, sind diese beiden Datenquellen prädestiniert für die Bildung von Korpora, die zusätzlich mit einer allgemeinen themenverwandten Stichprobe von Twitternutzern ergänzt werden können. Anhand dieser extrahierten Datenquellen lassen sich neben Gruppendynamiken und verschiedenen Rollenstrukturen vor allem Sprachgebräuche und Textcharakteristika im Rahmen korpuslinguistischer Analysen gewinnen, die hilfreich zur Beantwortung der Fragen sind, ob sich Personen, die einer bestimmten Ideologie zugeneigt sind, sich anhand ihres Sprachgebrauchs

identifizieren lassen und ob sich anhand dessen Indikatoren finden lassen, die besonders wirksam für eine Identifizierung ebendieser Akteure sind. Des Weiteren werden auch die Unterschiede und Gemeinsamkeiten zwischen den TG betrachtet, die durch die drei Korpora abgebildet wurden.

Dabei fällt im Rahmen der korpuslinguistischen Analysen auf, dass in Korpus C der Gebrauch positiv konnotierter Terme deutlich häufiger vorzufinden ist, als es in den beiden IS-verwandten Korpora der Fall war. Interessante Erkenntnisse liefert außerdem ein Plot der prozentualen relativen Termhäufigkeiten über die jeweiligen Magazine, der sehr genau die militärische Bedrohungslage und den Einsatz propagandistischer Motive des IS widerspiegelt. Auch die Clusteranalysen dienen hervorragend als Indikator zur Charakterisierung des Sprachgebrauchs und zur Detektion von ähnlichen Wortsequenzen, die besonders häufig zwischen Korpus A und Korpus B vorzufinden sind und den Unterschied zwischen gläubigen Muslimen und fundamentalistischen Extremisten verdeutlichen. Auch die Betrachtung von Kookkurrenzen veranschaulicht, dass in Korpus A und Korpus B die Bedeutung des Terms „allah“ höchst theologisch ausgelegt wird, während diese in Korpus C eher positive Eigenschaften zuordnen. Noch klarer sind die Unterschiede bei der Betrachtung des Terms „jihad“, der in Korpus A und B dem Sprachgebrauch nach als legitimes Mittel, in Korpus C jedoch als sehr negativ beschrieben wird. Graphisch nachvollziehbar ist dies in Form von Wordclouds noch deutlicher. Ein ebenso besonders wirksamer Indikator ist die Aufstellung von Keyword Lists, die besonders die Gemeinsamkeiten und Unterschiede von Korpus B und Korpus C in Bezug auf Korpus A hervorhebt. Dabei fällt im Vergleich zu Korpus A auf, dass in Korpus B die Sprache massiv gewaltverherrlichender ist, während sie in Korpus C deutlich positiver ausfällt. Dies macht besonders anschaulich, dass sich der IS in Form seiner Propagandamagazine ein eher theologisch professionelles Selbstbild auflegen will, während die Sympathisanten und Akteure des IS in sozialen

Netzwerken, die im Rahmen ihrer Online-Strategie genutzt werden, deutlich hemmungsloser und offener kommunizieren.

Die anschließend durchgeführten statistischen Signifikanztests führen demzufolge außerdem in neun von zehn Fällen zu dem Ergebnis, dass mit 99,9% Wahrscheinlichkeit davon ausgegangen werden kann, dass die Frequenzunterschiede von mehreren Termen zwischen den Korpora nicht zufällig, sondern absichtlich zustande gekommen sind. In lediglich einem Fall beträgt die Wahrscheinlichkeit nur 95%, was jedoch immer noch als signifikant betrachtet werden kann. Diese Signifikanztests beweisen also mathematisch die bereits gewonnenen Ergebnisse und sind eine sinnvolle Ergänzung zur Nutzung der korpuslinguistischen Analysen, um diese zu verifizieren. Ähnlich gilt dies auch für die Durchführung der Sentimentanalysen, die die gesammelten Ergebnisse stützt und deutlich macht, dass die Haltung und Nutzung des Sprachgebrauchs in Korpus A und B merklich negativer als in Korpus C behaftet ist.

Schlussendlich lässt sich also festhalten, dass Korpuslinguistische Analysen ein Werkzeug darstellen, mit dessen Hilfe sich in sicherheitspolitischem Kontext sehr wirksam gegen terroristische Aktivitäten vorgehen lässt und das sich sinnvoll mit Signifikanztests und SA ergänzen lässt. Dabei sind besonders Keyword Lists, Clusteranalysen und Kookkurrenzanalysen sinnvolle Indikatoren, anhand derer sich Personen und Sprachcharakteristika vermessen lassen, die einer terroristischen Ideologie zugeneigt sind. Die wichtige Voraussetzung, dass diese Verfahren Unterschiede zu nichtterroristischen Gruppen dabei hinreichend berücksichtigen, um keine False Positives zu erstellen, ist dabei durchgehend gegeben. Die Unterschiede der beiden Korpora, die der radikalen Gruppe zugeordnet sind, zeigen während der Analysen durchgehend derart große Unterschiede zum Korpus der allgemeinen Twitternutzer im Sprachgebrauch, dass abschließend festgehalten werden kann, dass zwischen der Allgemeinheit und

fundamentalistischen Personen große Unterschiede in der Sprachcharakteristik bestehen, die sich mithilfe korpuslinguistischer Analyse skalieren und vergleichen lassen und sich damit als wirksames Mittel zur Terrorismusbekämpfung nutzen lassen.

6 Literaturverzeichnis

Wissenschaftliche Literatur

- Ashcroft, M., Fisher, A., Kaati, L., Omer, E., & Prucha, N. (2015). Detecting Jihadist Messages on Twitter. *2015 European Intelligence and Security Informatics Conference, Manchester, 2015* (S.161-164). o. O., o. V..
- Athira, U., & Thampi, S. M. (2015). Authorship Analysis: Techniques and Challenges. In M. Dawson (Hrsg.), *New Threats and Countermeasures in Digital Crime and Cyber Terrorism* (S. 173-194). United States of America: IGI Global.
- Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Social Science and Public Policy*, 54(1), 138-149.
- Benigni, M. C., Joseph, K., & Carley, K. M. (2017). Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter. *PloS ONE*, 12(12), 1-23.
- Benigni, M. C., Joseph, K., & Carley, K. M. (2018). Mining online communities to inform strategic messaging: practical methods to identify community-level insights. *Computational and Mathematical Organization Theory*, 24(1), 224-242.
- Bredack, J., & Lepsky, K. (2014). Automatische Extraktion von Fachterminologie aus Volltexten. *ABI Technik*, 34(1), 2-8.
- Brynielsson, J., Horndahl, A., Johansson, F., Kaati, L., Mårtenson, C., & Svenson, P. (2013). Harvesting and analysis of weak signals for detecting lone wolf terrorists. *Security Informatics*, 2(1), 11.
- Cohen, K., Johansson, F., Kaati, L., & Mork, C. (2014). Detecting Linguistic Markers for Radical Violence in Social Media. *Terrorism and Political Violence*, 26(1), 246-256.
- Cohen, S. J., Kruglanski, A., Gelfand, M. J., Webber, D., & Gunaratna, R. (2018). Al-Qaeda's Propaganda Decoded: A Psycholinguistic System for Detecting Variations in Terrorism Ideology. *Terrorism and Political Violence*, 30(1), 142-171.
- El Ghamari, M. (2017). Pro-Daesh Jihadist Propaganda: A Study of Social Media and Video Games. *Security and Defence Quarterly*, 14(1), 69-90.
- Ghajar-Khosravi, S., Kwantes, P., Derbentseva, N., & Huey, L. (2016). Quantifying Salient Concepts Discussed in Social Media Content: An Analysis of Tweets Posted by ISIS Fangirls. *Journal of Terrorism Research*, 7(2), 79-90.
- Grothe, M. (2016). Computerlinguistisches Echtzeitmonitoring im Internet: Digital Listening. *Information – Wissenschaft & Praxis*, 67(2-3), 117-122.

- Hamblet, M. (2017). Jihad in the West: The Islamic State's Virtual Caliphate. *Middle East Quarterly*, 24(4), 1-8.
- Hirschmann, H. (2019). *Korpuslinguistik. Eine Einführung*. o. O.: J. B. Metzler.
- Huey, L., Inch, R., & Peladeau, H. (2019). "@ me if you need shoutout": Exploring Women's Roles in Islamic State Twitter Networks. *Studies in Conflict & Terrorism*, 42(5), 445- 463.
- Ingram, H. J. (2016). An analysis of Islamic State's Dabiq magazine. *Australian Journal of Political Science*, 51(3), 458-477.
- Ingram, H. J. (2018). Islamic State's English-language magazines, 2014-2017: Trends & implications for CT-CVE strategic communications. *The International Centre for Counter-Terrorism – The Hague* 8(15), 1-46.
- Jacoby, T. (2019). Islam and the Islamic State's Magazine, Dabiq. *Politics and Religion*, 12(1), 32-54.
- Khwaileh, T., & Khuwaileh, A. (2012). The semantic confusing connotation of terrorism: Terminologies generated in Arabic and used in English. *Babel*, 58(1), 19-30.
- Kienitz-Vollmer, B., & Reichardt, J. (1986). Bestimmung von Mehrwortgruppen mithilfe des Begrenzerverfahrens. In G. Lustig (Hrsg.), *Automatische Indexierung zwischen Forschung und Anwendung* (S. 18-30). Hildesheim, Zürich, New York: Olms.
- Klausen, J., Marks, C. E., & Zaman, T. (2016). Finding Online Extremists in Social Networks. *Operations Research*, 66(4), 957-976.
- Lustig, G. (1986). Verfahren zur Gewinnung von Wörterbuchdaten. In G. Lustig (Hrsg.), *Automatische Indexierung zwischen Forschung und Anwendung* (S. 13-17). Hildesheim, Zürich, New York: Olms.
- Macnair, L., & Frank, R. (2018a). Changes and stabilities in the language of Islamic state magazines: a sentiment analysis. *Dynamics of Asymmetric Conflict*, 11(2), 109-120.
- Macnair, L., & Frank, R. (2018b). The mediums and the messages: exploring the language of Islamic State through sentiment analysis. *Critical Studies on Terrorism*, 11(3), 438- 457.
- Mansour, S. (2018). Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. In C. H. Dagli (Hrsg.), *Cyber Physical Systems and Deep Learning, Chicago, Illinois November 5-7, 2018* (S. 95-103). o. O.: Elsevier.
- Marcellino, W. M., Cragin, K., Mendelsohn, J., Cady, A. M., Magnuson, M., & Reedy, K. (2017). Measuring the Popular Resonance of Daesh's Propaganda. *Journal of Strategic Security*, 10(1), 32-52.

- Mehl, M. R., & Gill, A. J. (2010). Automatic Text Analysis. In S. D. Gosling, & J. A. Johnson (Hrsg.), *Advanced Methods for Conducting Online Behavioral Research* (S. 109-127). Washington, DC: American Psychological Association.
- Melki, J., & Jabado, M. (2016). Mediated Public Diplomacy of the Islamic State in Iraq and Syria: The Synergistic Use of Terrorism, Social Media and Branding. *Media and Communication*, 4(2), 92-103.
- Naab, T. K., & Sehl, A. (2014). Inhaltsanalytische Untersuchungen von User-Generated-Content-Angeboten: Eine Bestandsaufnahme zur Anwendung der Methode. In K. Sommer, M. Wettstein, W. Wirth, & J. Matthes (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S. 127-144). Köln: von Halem.
- Novenario, C. M. I. (2016). Differentiating Al Qaeda and the Islamic State Through Strategies Publicized in Jihadist Magazines. *Studies in Conflict & Terrorism*, 39(11), 953-967.
- Pearson, E. (2018). Online as the New Frontline: Affect, Gender, and ISIS-Take-Down on Social Media. *Studies in Conflict & Terrorism*, 41(11), 850-874.
- Piskorski, J., & Yangarber, R. (2013). Information Extraction: Past, Present and Future. In T. Poibeau, H. Saggion, J. Piskorski, & R. Yangarber (Hrsg.), *Multi-source, Multilingual Information Extraction and Summarization* (S. 23-52). Berlin, Heidelberg: Springer.
- Richards, I. (2016a). "Flexible" capital accumulation in Islamic State social media. *Critical Studies on Terrorism*, 9(2), 205-225.
- Richards, I. (2016b). "The Spirit of Terrorism" in Islamic State Media. *International Journal of Baudrillard Studies*, 13(2), 1-31.
- Richey, M. K., & Binz, M. (2015). Open Source Collection Methods for Identifying Radical Extremists Using Social Media. *International Journal of Intelligence and CounterIntelligence*, 28(2), 347-364.
- Saggion, H., & Poibeau, T. (2013). Automatic Text Summarization: Past, Present and Future. In T. Poibeau, H. Saggion, J. Piskorski, & R. Yangarber (Hrsg.), *Multi-source, Multilingual Information Extraction and Summarization* (S. 3-22). Berlin, Heidelberg: Springer.
- Scharkow, M. (2012). *Automatische Inhaltsanalyse und maschinelles Lernen*. Berlin: Epubli.
- Siegel, M., Deuschle, J., Lenze, B., Petrovic, M., & Starker, S. (2017). Automatische Erkennung von politischen Trends mit Twitter – brauchen wir Meinungsumfragen noch?. *Information – Wissenschaft & Praxis*, 68(1), 67-74.

- Spitzberg, B. H., & Gawron, J. M. (2016). Toward Online Linguistic Surveillance of Threatening Messages. *Journal of Digital Forensics, Security and Law*, 11(3), 43-78.
- Stern, J., & Berger, J. M. (2015). *ISIS: the state of terror*. New York: Ecco.
- Stock, W. G., & Stock, M. (2008). *Wissensrepräsentation: Informationen auswerten und bereitstellen*. München: Oldenbourg.
- Striegher, J.-L. (2013). Early detection of the lone wolf: advancement of counter-terrorism investigations with an absence or abundance of information and intelligence. *Journal of Policing, Intelligence and Counter Terrorism*, 8(1), 35-53.
- Şuşnea, E. (2018). A Real-Time Social Media Monitoring System as an Open Source Intelligence (Osint) Platform for Early Warning in Crisis Situations. *International conference KNOWLEDGE-BASED ORGANIZATION*, 24(2) (S. 427-431). o. O.: Sciendo.
- Veilleux-Lepage, Y. (2016). Paradigmatic Shifts in Jihadism in Cyberspace: The Emerging Role of Unaffiliated Sympathizers in Islamic State's Social Media Strategy. *Journal of Terrorism Research*, 7(1), 36-51.
- Weimann, G. J. (2019). Competition and Innovation in a Hostile Environment: How Jabhat Al-Nusra and Islamic State Moved to Twitter in 2013-2014. *Studies in Conflict & Terrorism*, 42(1-2), 25-42.
- Wettstein, M. (2014). „Best of both worlds”: Die halbautomatische Inhaltsanalyse. In K. Sommer, M. Wettstein, W. Wirth, & J. Matthes (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S. 16-39). Köln: von Halem.

Genutzte Online-Ressourcen und Tutorials

- Clarion Project (o. D.a). *About Clarion Project*. Abgerufen 16.06.2020, von <https://clarionproject.org/about-us/>
- Clarion Project (o. D.b). *Islamic State's (ISIS, ISIL) Horrific Magazine*. Abgerufen 16.06.2020, von <https://clarionproject.org/islamic-state-isis-isil-propaganda-magazine-dabiq-50/>
- Informatik Aktuell (2020). *Herangehensweisen und Konzepte der Funktionalen Programmierung mit Scala*. Abgerufen 18.06.2020, von <https://www.informatik-aktuell.de/entwicklung/programmiersprachen/funktionale-programmierung-mit-scala-herangehensweisen-und-konzepte.html>
- Rüdiger, J. O. (2015). *CorpusExplorer v2.0. Benutzerhandbuch 2015-2018*. Abgerufen 30.06.2020, von <http://www.bitcutstudios.com/products/CorpusExplorer/help-2/#>

Scala (o. D.). *Ein Scala Tutorial für Java Programmierer*. Abgerufen 18.06.2020, von <https://docs.scala-lang.org/de/tutorials/scala-for-java-programmers.html>

Scala Exercises (o. D.). *The Path to Enlightenment*. Abgerufen 18.06.2020, von <https://www.scala-exercises.org/>

Schäfer, S. (o. D.). *Scala Tutorial. Deutsches Tutorial für Umsteiger imperativer Programmiersprachen*. Abgerufen 18.06.2020, von <https://scalatutorial.wordpress.com/>

Tutorialspoint (o. D.). *Scala Tutorial*. Abgerufen 18.06.2020, von <https://www.tutorialspoint.com/scala/index.htm>

Genutzte Software

Anthony, L. (2019). AntConc (3.5.8) [Software]. Abgerufen 30.06.2020, von <http://www.laurenceanthony.net/software/antconc/>

Apache Software Foundation (2020). Apache Spark (2.4.5) [Software]. Abgerufen 10.04.2020, von <https://spark.apache.org/>

Brooker, P., Barnett, J., & Cribbin, T. (2016). Chorus Tweetcatcher Desktop Edition (1.3.1) [Software]. Abgerufen 04.05.2020, von <http://chorusanalytics.co.uk/>

CyberEmotions (o. D.). SentiStrength (2.3.6977.32036) [Software]. Abgerufen 21.07.2020, von <http://sentistrength.wlv.ac.uk/>

Rüdiger, J. O. (2018). CorpusExplorer (2.0.) [Software]. Abgerufen 30.06.2020, von <http://www.corpusexplorer.de>

Anhang

Anhang 1: Scala-Programmcode zur Datenextraktion mit Apache Spark 2.4.5

```
import scala.util.Try
import scala.util.parsing.json.JSON

val twitter_streams_root = "file:///data/twitter/streams/"
val streamName = "prometheus/"
val day = "statuses.log.2013-01-31.gz/"
val tweetlinesRDD = sc.textFile(twitter_streams_root + streamName + day)
val tweetAsJsonRDD = tweetlinesRDD.filter{ line => line.contains("#AllEyesOnISIS")
}.map { line => JSON.parseFull(line)
}.map { _.getOrElse("":Any)
}.map { tweet =>
  Try(tweet.asInstanceOf[Map[String,Any]]).getOrElse(Map("":->""))
}.map{t => t.get("text")}
val t = tweetAsJsonRDD.take(15)
t.foreach(println)
```


Anhang 2: Genutzte Suchanfragen zur Erstellung von Korpus B

Profile	Hashtags
@_lady_ghuraba_	#al-furqan
@_mrsl337	#al-hayat
@445468a7e3fc45c	#alleyesonisis
@613780	#amessagefromisistous
@aarishmajeed	#baqiya
@abdisamed3	#baqiyya
@abdulnagi313	#calamaitywillbefallus
@abou_cha_bouche	#ei
@abuhudayr03	#helparakan
@abusulayman321	#iloveisis
@ak47zneedlove	#ilovejihad
@alfurqan2013	#is
@aljabarti28	#islamicrevival
@amreekiwitness	#islamicstate
@an_najdee	#islamicstate
@ansaralummah49	#islamicstate
@aqidahhaqq	#jihad
@backup_109_109	#jihadifitness
@banukombe	#kaffir
@burnningone	#kafir
@buruan8	#khalifarestored
@catseikh	#khilafah
@citizenkhil4f4	#khilfa
@e3tasimo	#kufar
@enkorela	#kuffar
@hashtag_isis	#kuffr
@hemm_agency	#kufr
@isilcats	#littlemujahid
@jabhtannusrah	#mujahidin
@jbhtannusrah	#murtadd
@khalidbinalwale	#sevensseven
@men9174	#sykespicotover
@mghol1122	#teamhijabi
@minbar_s	#thefridayofsupportingisis
@musabgharieb13	#thought_of_a_lone_lion
@musabgharieb18	#women_under_isis
@nnews	#women_under_isis.
@omarabdullah_00	
@omgnotagain313	
@ouaicheu	

<p> @ouaisheu @oum1337 @oumhu554in @oumhu554inbrit @oumhu55ain @oumhussaln @oumhussain @poorslave_3 @profomar0 @qistheword_ @raqqa98 @salafi_jihadist @saraeuk @shamiwitness @stillukhtmaryam @tahrirsy @the_ansar_4_isis @umm__maryam @ummhu55ain2 @ummhussain9ll @ummlayth @urwatulwuthqa_ allahumawlakum @usamah_m @w_raqqa @wa3tasimu @with_baghdadi @yaqub_london </p>	
---	--

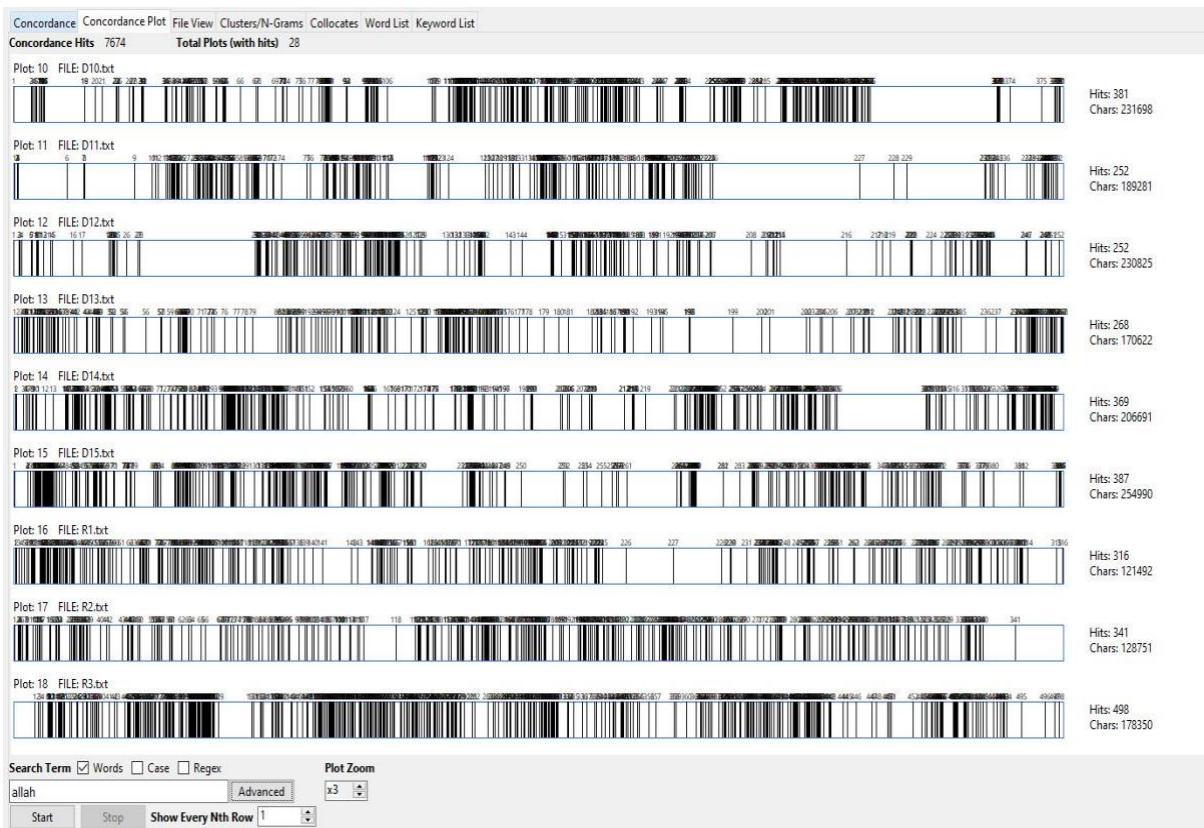
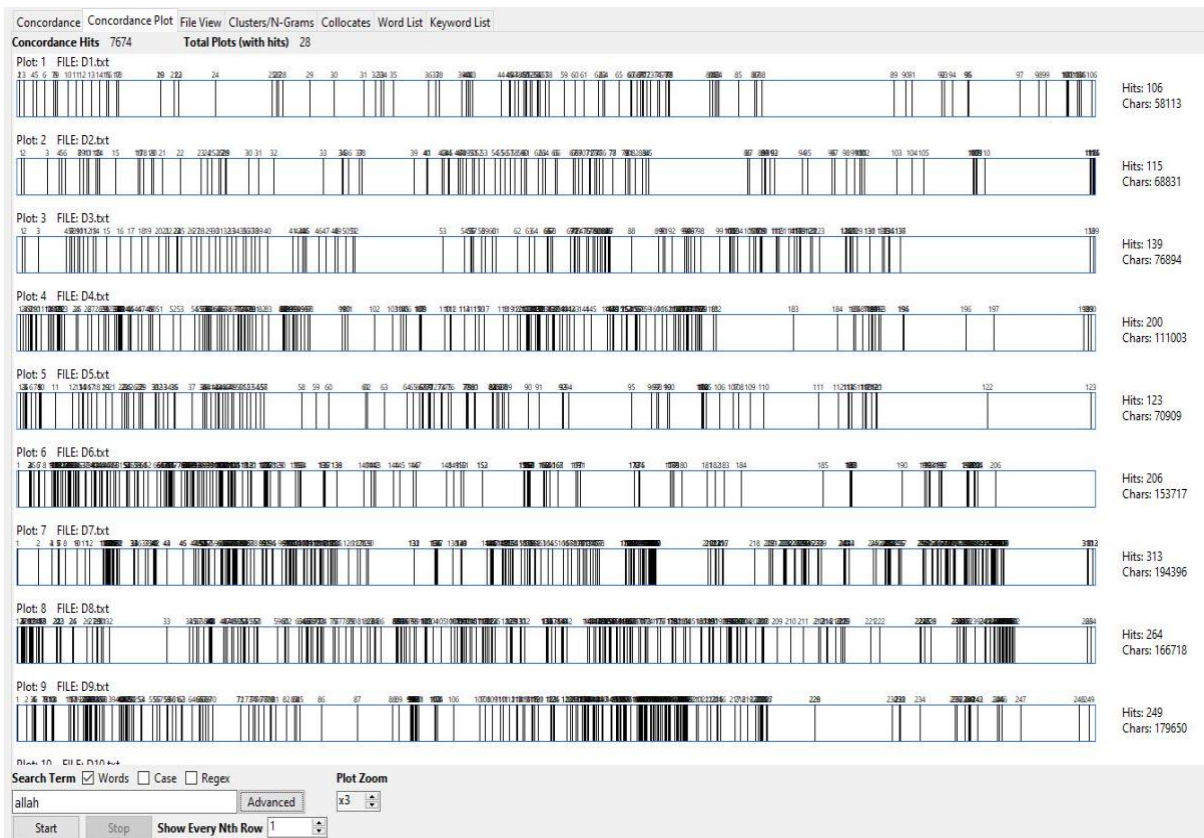
Anhang 3: Genutzte Suchanfragen zur Erstellung von Korpus C

Profile	Hashtags und Terms
@assimalhakeem @diyaaul_islam @islamicfreedom @islamicstrength @islamic tongue @joburgmuslim @muslimsmatters @retweetprayers @thenoblequran @topislamicnet	#allah #deen #islam #jihad #jumah #jummah #muslim #muslimsaroundtheworld #naat #nasheed #prophetmuhammad #quraan #quran #taij #versesfromthequran caliphate isil isis islamic and state mujahid

Anhang 4: Stoppwortliste zur Top-20 Termextraktion und Regular Expressions

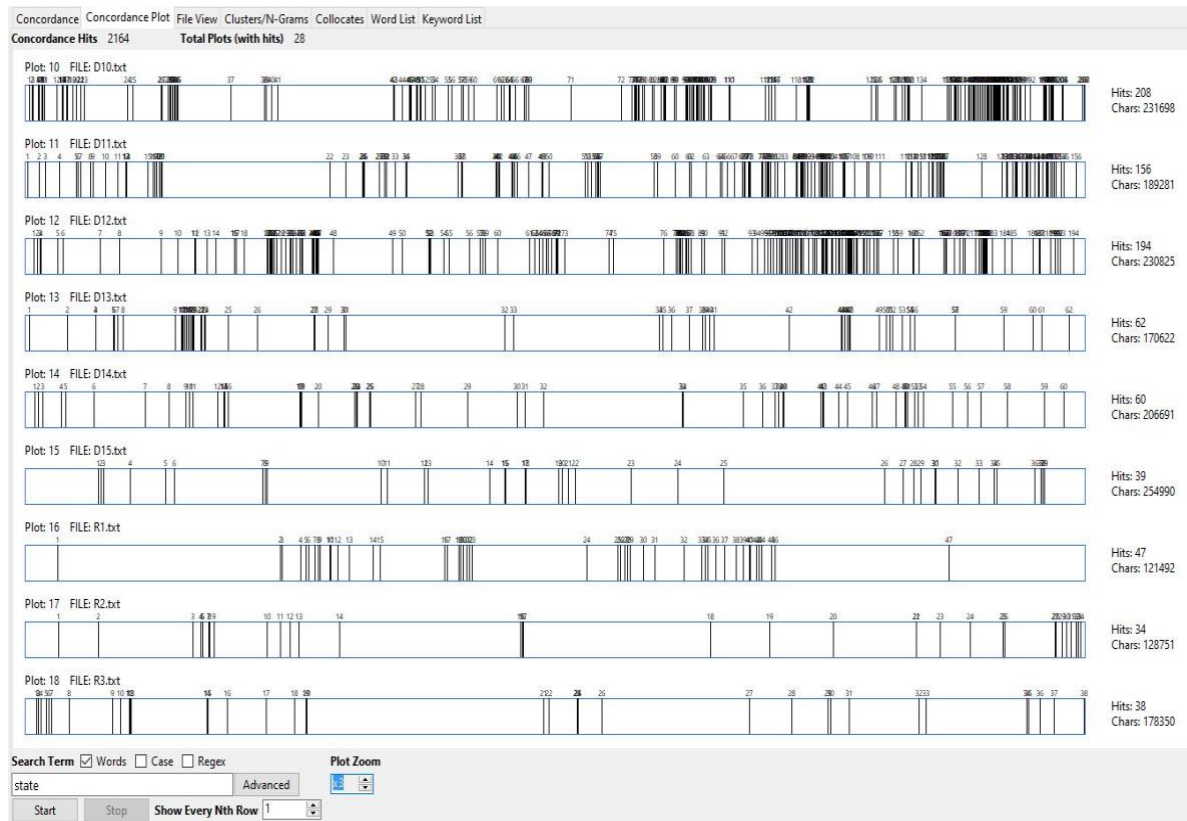
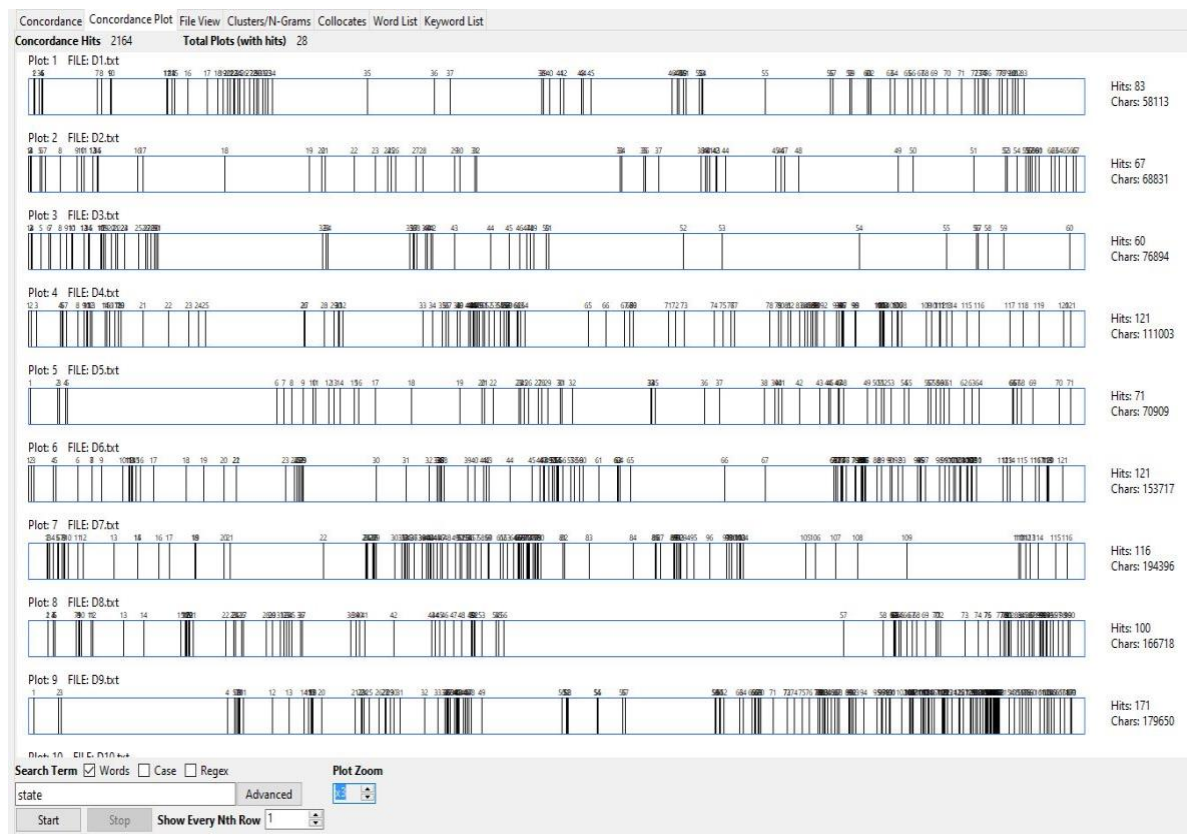
Korpus A	Korpus B	Korpus C
the, of, and, to, in, a, is, that, for, he, they, their, it, his, as, with, them, from, you, by, al, not, who, was, this, on, s, are, be, said, him, so, will, have, i, then, were, one, we, those, which, an, would, or, what, if, has, against, ibn, all, its, at, but, your, when, upon, do, there, after, ah, had, no, these, also, us, being, out, o, than, more, indeed, any, even, some, until, such, my, into, while, only, other, our, me, did, been, before, over, two, may, many, about, except, say, g, her, because, among, most	the, of, to, in, is, and, a, for, you, on, are, i, s, this, by, it, that, they, with, al, not, we, from, sevenseven, them, an, if, u, be, amessagefromisistous, tcot, will, t, who, all, what, have, was, their, as, us, he, your, about, so, at, but, no, has, one, my, out, alleyesonisis, do, our, or	the, to, of, is, you, and, a, in, for, i, your, it, are, that, s, with, have, on, not, this, be, he, who, me, t, by, from, if, will, they, them, we, us, n, when, my, all, as, one, an, so, has, nn, what, his, no, but, can, was, at, our, than, do, him, make, never, their, don, ya, only, because, more, most, about, how, or, there, remember, side, may, al, other, know, say, before, every, even, m, just, those, out, last, best, then, o, like, been, today, which, two, up, against, go
RE Korpus A	RE Korpus B	RE Korpus C
^p ^p → [_] ^p [.] ^{\$} → löschen ^[0-9] → löschen [_] ^[b-z] [_] → löschen [—] → [_]	Some(→ löschen) + (→ löschen \@*?_ → löschen → [.] _____ → löschen http*?_ → löschen RT_ → löschen [.] → [^] ^p &*?_ → löschen [^] ^p [^] ^p → [^] ^p [^] ^p [_] → [^] ^p	???_???_?:^t → löschen [^] ^p → löschen [^] ^t 12*_?*t... → löschen http?*_ → löschen RT_ → löschen \@*?_ → löschen [^] ^p → S1904 [^] ^p S1904?* [^] ^t → [^] ^p [...] → löschen

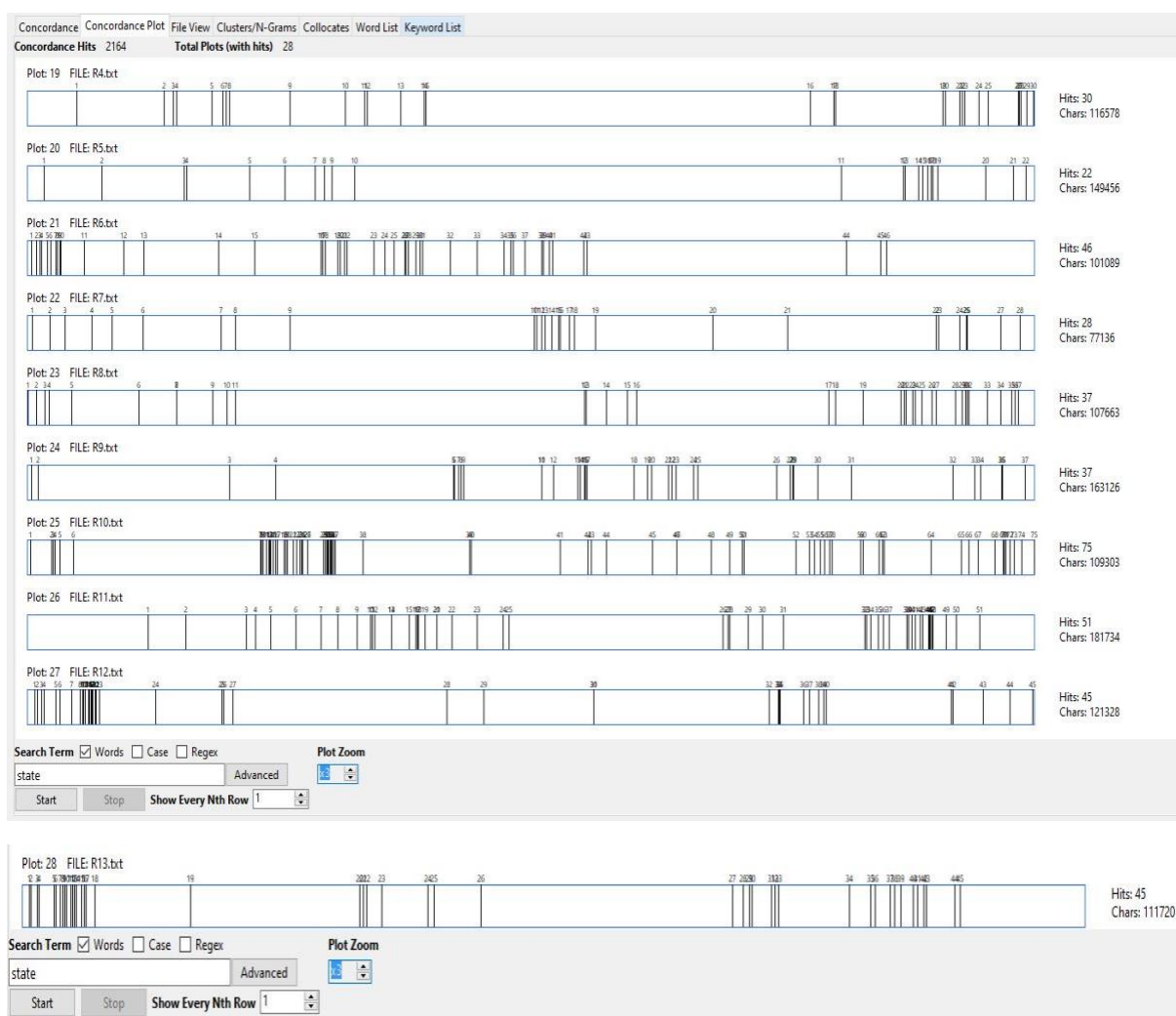
Anhang 5: Konkordanz-Plots am Beispiel der IS-Magazine mit Term „allah“





Anhang 6: Konkordanz-Plots am Beispiel der IS-Magazine mit Term „state“





Anhang 7: Clusteranalysedaten mit Bezugsterm „allah“ und Frequenz

Korpus A	Korpus B	Korpus C
460 allah's messenger	16 god but allah	1352 be to allah
258 by allah's	15 allah, grant victory	1316 allah, even if
209 allah and his	15 allah, grant victory to	1316 allah, even if you
190 allah and his messenger	15 no god but allah	1316 allah, even if you have
184 of allah's	14 o allah, grant	1316 back to #allah
170 of allah and	14 o allah, grant victory	1316 back to #allah, even
169 cause of allah	14 o allah, grant victory to	1316 back to #allah, even if
169 the cause of allah	13 messenger of allah	1316 go back to #allah
158 allah's messenger g	12 allah muhammad is	1316 go back to #allah, even
137 allah's permission	12 allah muhammad is the	1316 to #allah, even
131 messenger of allah	12 allah muhammad is the	1316 to #allah, even if
119 and allah is	messenger	1316 to #allah, even if you
110 for allah's	12 but allah muhammad	1212 allah, please make
106 may allah accept	12 but allah muhammad is	1212 allah, please make
105 due to allah	12 but allah muhammad is	tomorrow
103 by allah's permission	the	1212 allah, please make
100 allah's grace	12 god but allah muhammad	tomorrow brighter
100 enemies of allah	12 god but allah muhammad	1212 dear allah, please
100 for the cause of allah	is	1212 dear allah, please make
97 is due to allah	12 is the messenger of allah	1212 dear allah, please make
97 praise is due to allah	12 messenger of allah #isis	tomorrow
95 by allah's grace	12 of allah #isis	1054 to allah is
93 the enemies of allah	12 the messenger of allah	1036 allah give me
89 allah c said	12 the messenger of allah	1036 ya allah give
89 allah's messenger	#isis	1036 ya allah give me
sallallāhu	11 allah #isis #is	993 times go back to #allah
89 allah's messenger	11 messenger of allah #isis	878 allah forgive me
sallallāhu 'alayhi	#is	782 to allah we
84 allah's cause	11 no god but allah	770 back to allah
84 the messenger of allah	muhammad	762 brought you back to
82 of allah's messenger	11 of allah #isis #is	allah
77 with allah's	10 may allah accept	762 you back to allah
75 allah and the	8 allah's sight	729 allah is the
70 ask allah to	8 allah's sight are	720 allah those that
69 for allah's cause	8 allah's sight are those	720 allah those that have
68 in allah and	8 allah, grant victory to the	720 allah those that have
63 allah (ta'ālā	8 animals in allah	fallen
63 in the cause of allah	8 animals in allah's	718 ya allah those
63 that allah's	8 animals in allah's sight	717 ya allah those that
63 we ask allah	8 in allah's	717 ya allah those that have
62 allah accept him	8 in allah's sight	702 ya allah forgive

62 may allah accept him 60 to allah and 60 to allah, the 59 sake of allah 58 for the sake of allah 58 the sake of allah 56 of allah, and 55 allah, the lord 55 indeed, allah is 54 allah, the lord of 54 that allah's messenger 51 statement of allah 50 other than allah 50 to allah, the lord	8 in allah's sight are 8 may allah protect 8 of animals in allah 8 of animals in allah's 8 vilest of animals in allah 7 allah, grant victory to t 7 sake of allah 7 the sake of allah 6 allah !!! grant victory 6 allah !!! grant victory to 6 allah !!! grant victory to the 6 allah nor the 6 allah nor the last 6 allah nor the last day 6 allah the revealer 6 allah the revealer of 6 allah the revealer of the 6 allah vp hamid 6 allah vp hamid ansari 6 allah vp hamid ansari refuses 6 anyone but allah	701 ya allah forgive me 689 allah we belong 689 to allah we belong 688 allah has a 680 to allah, we 676 ya allah, forgive 675 praise be to allah 659 insha'allah for 658 alhamdulillah for today, insha'allah 658 allah for tomorrow 658 for today, insha'allah 658 for today, insha'allah for 658 insha'allah for tomorrow 658 today, insha'allah 658 today, insha'allah for 658 today, insha'allah for tomorrow 648 allah is a 645 allah for being 640 may allah bless 623 allah knows what 603 trust in allah
--	--	--

Anhang 8: Clusteranalysedaten bezogen auf n-Gramme und Frequenz

Korpus A	Korpus B	Korpus C
1559 the islamic state	136 what is jihad	5916 islamic state of
597 soldiers of the	133 so what is	3172 a history of
497 of the islamic	133 solution to the	3099 they led an
495 the soldiers of	132 so what is jihad	3098 have a history
475 the soldiers of the	131 immediate solution to	3098 have a history of
460 allah s messenger	131 immediate solution to	3098 led an armed
445 of the islamic state	the	3098 they led an armed
382 alayhi wa sallam	131 jihad watch this	3097 an armed militancy
369 of the khilafah	131 jihad watch this video	3097 an armed militancy to
366 sallallāhu alayhi wa	131 jihad watch this video	3097 armed militancy to
364 sallallāhu alayhi wa	to	3097 led an armed
sallam	131 the immediate solution	militancy
363 there is no	131 the immediate solution	3097 led an armed
346 the th of	to	militancy to
345 the people of	131 the immediate solution	3097 they led an armed
344 the city of	to the	militancy
324 on the th	131 this video to	3096 a history of religious
321 on the th of	131 this video to found	3096 a history of religious
309 a number of	131 video to found	service
304 as well as	131 watch this video	3096 have a history of
299 one of the	131 watch this video to	religious
286 of the khilāfah	131 watch this video to	3096 history of religious
279 of allah and	found	3096 history of religious
273 soldiers of the khilafah	130 immediate solution to	service
258 by allah s	the radicalisation	3096 nnin they led
250 in order to	130 is jihad watch	3096 nnin they led an
244 of the muslims	130 is jihad watch this	3096 of religious service
231 the soldiers of the	130 is jihad watch this	3096 tablighis have a
khilafah	video	3096 tablighis have a
215 allah and his	130 opposes terrorism so	history
215 the prophet g	130 opposes terrorism so	3096 tablighis have a
212 reported by al	what	history of
208 against the islamic	130 opposes terrorism so	3095 history of religious
203 against the islamic	what is	service nnin
state	130 so what is jihad watch	3095 nnin they led an
203 the sharī ah	130 solution to the	armed
201 and his messenger	radicalisation	3095 of religious service
196 the cause of	130 solution to the	nnin
193 war against the	radicalisation of	3095 religious service nnin
190 allah and his	130 terrorism so what	3094 of religious service
messenger	130 terrorism so what is	nnin they

185 some of the	130 terrorism so what is	3094 religious service nnin
184 in the city	jihad	they
184 of allah s	130 the radicalisation of	3094 religious service nnin
182 of the prophet	130 this video to found out	they led
180 the day of	130 to found out	3094 service nnin they
177 the religion of	130 to the radicalisation	3094 service nnin they led
170 cause of allah	130 to the radicalisation of	3094 service nnin they led
170 the cause of allah	130 video to found out	an
167 for the sake	130 what is jihad watch	3093 yes tablighis have
167 for the sake of	130 what is jihad watch	3093 yes tablighis have a
167 the sake of	this	3093 yes tablighis have a
159 in addition to	129 radicalisation of	history
159 in the city of	muslims	3092 side by side
158 allah s messenger g	129 the radicalisation of	3088 a barrier so
158 s messenger g	muslims	3088 a barrier so neither
153 the area of	129 to the radicalisation of	3088 a barrier so neither of
152 the lands of	muslims	3088 barrier so neither
149 as for the	128 the islamic state	3088 barrier so neither of
147 due to the	127 in the west	3088 barrier so neither of
147 of those who	127 islam opposes	them
146 and do not	terrorism	3088 between them is
145 al qā idah	127 islam opposes	3088 between them is a
141 by al bukhari	terrorism so	3088 between them is a
140 the jawlānī front	127 islam opposes	barrier
139 the muslims and	terrorism so what	3088 by side between
139 the shari ah	127 jihad and the	3088 by side between them
139 this is the	126 and the immediate	3088 by side between them
138 islamic state and	126 and the immediate	is
138 it is not	solution	3088 is a barrier
137 allah s permission		3088 is a barrier so
		3088 is a barrier so neither
		3088 meeting side by
		3088 meeting side by side

Anhang 9: Kookkurrenzen in Korpus A mit Frequenz

Term	Kookkurrenz	Frequenz
allah	messenger	916
	may	453
	cause	322
	praise	184
	permission	174
	slaves	145
	grace	142
	creation	130
	revealed	129
	ta'ālā	87
	belongs	78
	grant	75
	besides	53
	imrān	44
	azza	35
jihad	waging	53
	wage	46
	wealth	30
	tawhid	24
	kuffar	19
	obligation	14
	tawaghit	14
	ribat	12
	rumiyah	11
	shahadah	9
	waged	9
	soul	9
	muwahhidin	9
	siyahah	9
	trees	8

Anhang 10: Kookkurrenzen in Korpus B mit Frequenz

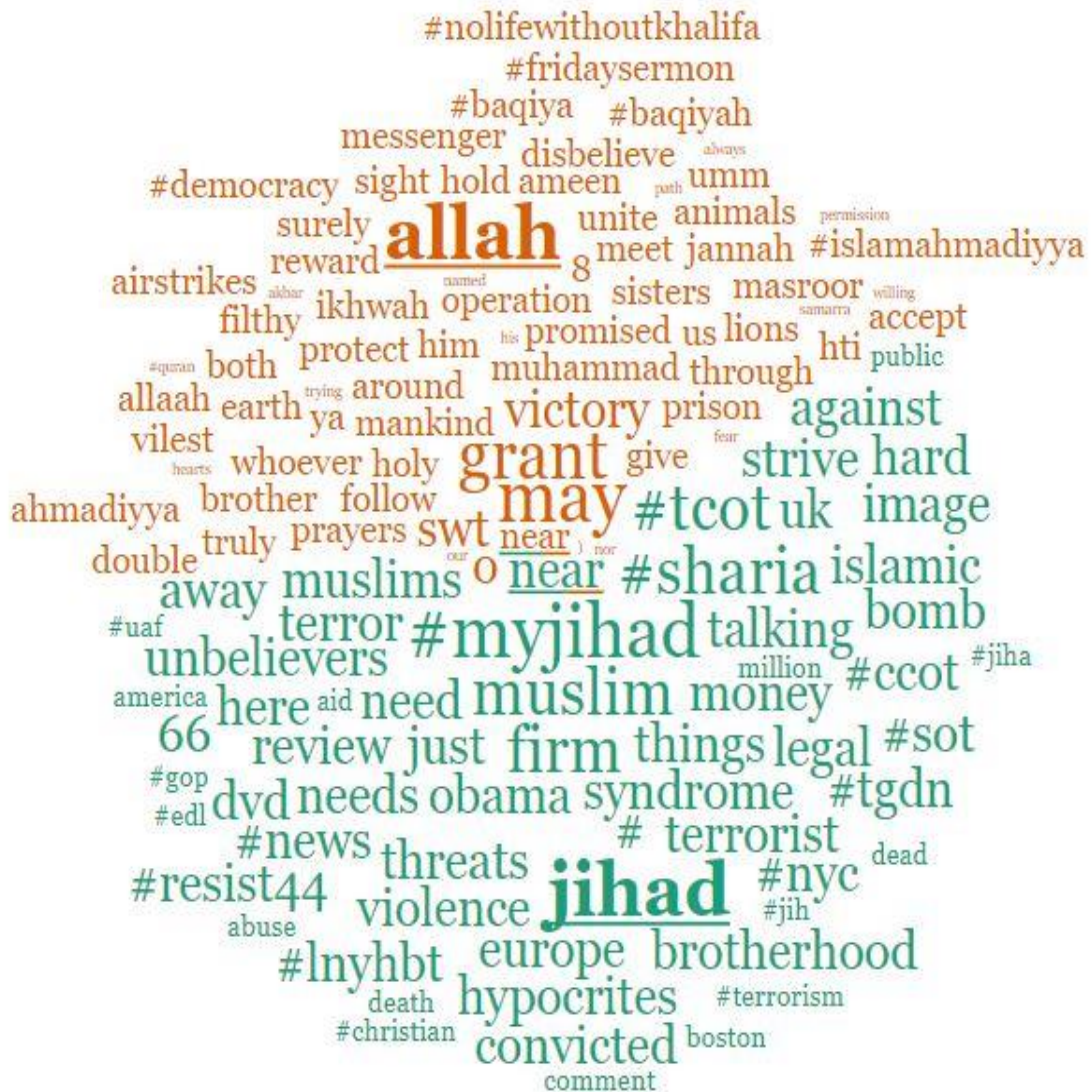
Term	Kookkurrenz	Frequenz
allah	may	44
	give	21
	grant	21
	o	20
	victory	19
	#islam	19
	follow	18
	swt	15
	muhammad	14
	protect	14
	brother	12
	ya	12
	#muslims	11
	sisters	10
	#muslim	10
jihad	#tcot	163
	sharia	116
	#jihadists	79
	#jihadist	60
	muslim	56
	#news	49
	#sharia	46
	#terrorism	43
	comment	39
	video	38
	#syria	36
	islam	35
	islamic	35
	#tgdn	35
	terrorism	34

Anhang 11: Kookkurrenzen in Korpus C mit Frequenz

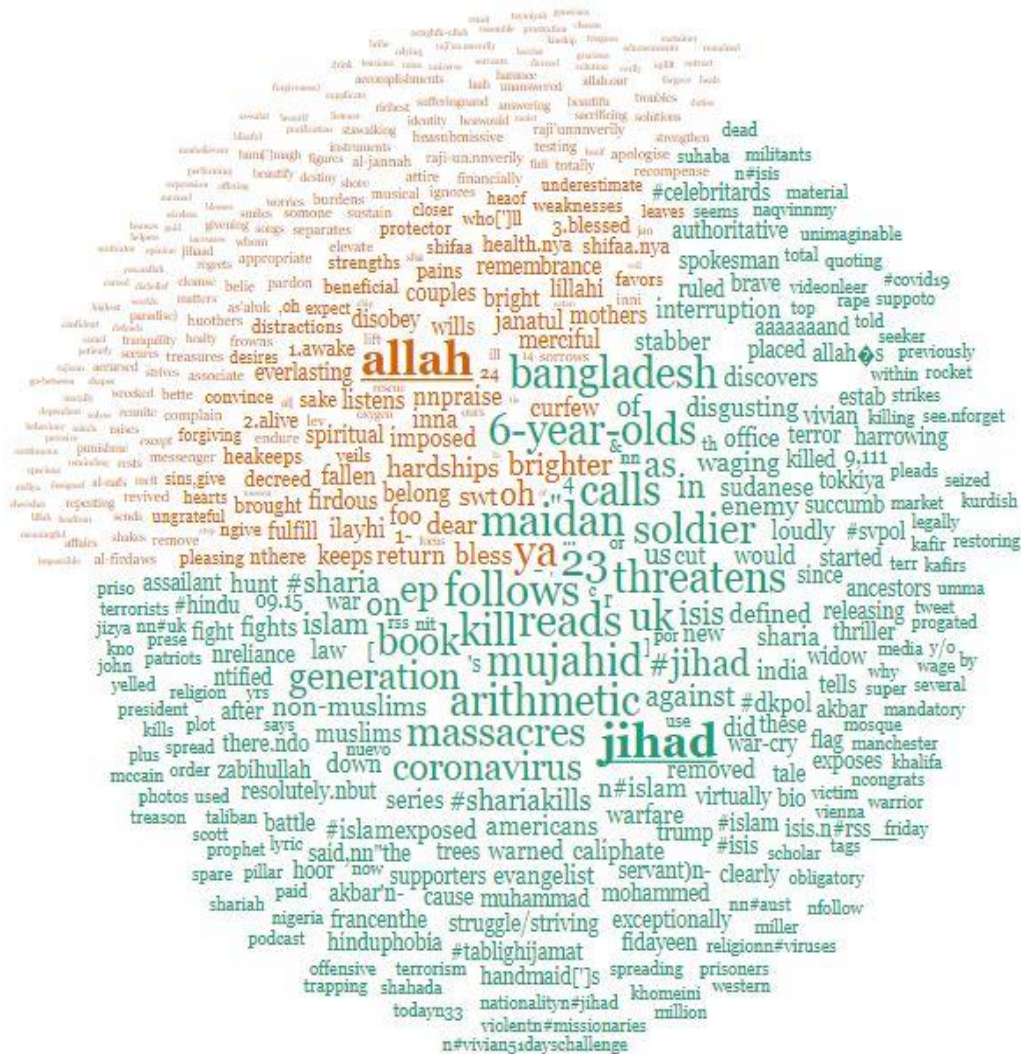
Term	Kookkurrenz	Frequenz
allah	ya	7226
	oh	2343
	bless	2118
	return	1657
	dear	1514
	brighter	1212
	keeps	1052
	brought	769
	hardships	765
	belong	686
	ill	685
	fallen	673
	merciful	672
	leaves	667
	curfew	589
jihad	mujahid	126
	kill	117
	maidan	116
	reads	116
	soldier	98
	uk	97
	massacres	95
	threatens	95
	muslims	44
	would	37
	war	35
	new	31
	non-muslims	28
	trump	28
	law	22



Anhang 13: Kookkurrenz-Wordcloud Korpus B



138



Anhang 15: Keyword List von Korpus B in Bezug zu Referenzkorpus A

Wort	Frequenz	Vorzeichen	Keyness
isis	2167	+	7627.46
islamicstate	1447	+	5501.32
jihad	1916	+	5144.63
myjihad	898	+	3414.09
caliphate	914	+	2894.71
sevenseven	750	+	2851.41
amessagefromisistous	687	+	2611.89
tcot	674	+	2562.47
syria	868	+	2329.74
isil	547	+	1839.29
islam	1377	+	1676.69
sharia	424	+	1606.21
alleyesonisis	343	+	1304.04
daesh	337	+	1235.94
terrorism	362	+	1140.85
video	296	+	949.34
u	338	+	877.65
iraq	607	+	865.79
watch	232	+	761.17
muslim	871	+	726.26
taliban	313	+	716.17
bokoharam	187	+	710.95
khilafah	568	+	695.88
jihadists	186	+	668.08
kings	211	+	650.79
jihadist	175	+	637.16
via	208	+	635.32
uk	182	+	597.33
news	318	+	593.54
israel	177	+	564.6
usa	149	+	538.45
mehdi	134	+	509.45
radicalisation	135	+	507.5
mi	139	+	500.5
london	148	+	484.02
t	625	+	475.09
caliph	134	+	470.95
twitter	126	+	456.58
alhayat	126	+	451.18
amp	115	+	437.22
immediate	133	+	432.51

assad	165	+	421.24
bombings	115	+	420.27
shia	137	+	420.27
islamdevleti	110	+	418.21
amaq	115	+	414.82
obama	197	+	413.9
breaking	144	+	408.7
jihadi	119	+	404.04
st	153	+	399.89
raqqa	115	+	399.04
solution	150	+	396.39
cont	104	+	395.4
opposes	134	+	392.32
maq	109	+	392.06
years	303	+	388.52
alhayatmediacenter	102	+	387.79
lol	102	+	387.79
furatmedia	101	+	383.99
ago	175	+	380.5
muawiyah	100	+	380.19
kobane	103	+	374.7
aleppo	130	+	365.75
ttp	96	+	364.98
umayyad	100	+	363.31
victims	115	+	347.56
htt	90	+	342.17
tgdn	90	+	342.17
vpn	90	+	342.17
terrorists	122	+	340.68
mujahideen	88	+	334.57
gaza	90	+	325.35
today	288	+	320.78
dont	84	+	319.36
is	3466	+	310.3
jihadis	81	+	307.95
get	173	+	301.88
service	113	+	283.39
edl	74	+	281.34
hey	75	+	279.42
islamophobia	71	+	269.93
don	209	+	269.17
saa	70	+	266.13
saudiarabia	69	+	262.33
ypg	80	+	262.15

remember	175	+	261.44
about	466	+	260.49
sex	68	+	258.53
terrorist	119	+	252.05
gohar	66	+	250.92
want	174	+	247.92
why	203	+	241.95
shahi	65	+	241.41
jn	64	+	237.61
mosque	64	+	237.61
british	121	+	237.5
now	272	+	236.97
gt	62	+	235.72
islamahmadiyya	62	+	235.72
promisedmessiah	62	+	235.72
alqaeda	61	+	231.91
amessagefromustois	61	+	231.91
sufyan	102	+	229.88
photo	66	+	224.05
rape	64	+	221.54
egypt	134	+	210.45
new	238	+	209.75
kathir	100	+	201.2
please	88	+	200.66
terror	123	+	190.81
media	183	+	188.94
damascus	77	+	188.09
pakistan	101	+	187.12
rebels	66	+	184.56
re	142	+	181.6
saudi	119	+	181.43
nigeria	67	+	180.6
m	127	+	178.22
arrested	68	+	173.46
pro	66	+	173.41
abi	105	+	170.77
de	63	+	169.99
anti	66	+	169.88
hamas	71	+	167.46
true	154	+	166.7
kurds	78	+	159.23
god	153	+	155.48
quran	153	+	154.25
mosul	107	+	148.51

women	247	+	148.29
really	82	+	147.37
india	70	+	146.66
anbar	63	+	146.41
forget	86	+	146.21
lost	111	+	145.84
i	1051	+	145.69
read	109	+	145.47
account	98	+	145.16
palestine	66	+	144.64
king	108	+	144.48
youth	86	+	139.98
flag	80	+	139.12
baghdadi	65	+	138.53
best	242	+	138.49
arabs	102	+	135.28
like	324	+	134.84
world	246	+	134.33
iran	135	+	133.25
says	139	+	129.23
hate	80	+	124.74
dar	103	+	123.96
muslims	649	+	122.88
hell	62	+	118.53
ll	61	+	117.84
still	124	+	116.08
stop	104	+	112.51
history	138	+	109.34
us	502	+	107.34
found	170	+	105.91
think	139	+	105.58
bomb	63	+	101.66
rumiyah	102	+	98.05
up	251	+	97.49
human	83	+	96.33
west	193	+	95.81
got	68	+	95.19
lives	108	+	94.18
christian	83	+	89.54
love	131	+	88.02
kurdish	76	+	86.36
dead	96	+	84.07
police	103	+	79.11
need	109	+	78.65

coming	79	+	77.89
old	80	+	77.73
off	129	+	76.64
turkey	103	+	76.33
peace	117	+	74.61
too	79	+	74.38
going	94	+	72.97
n	90	+	70.89
look	82	+	70.86
can	291	+	68.65
here	141	+	67.65
full	62	+	61.55
my	359	+	61
live	79	+	57.17
next	62	+	56.91
how	275	+	56.37
year	111	+	56.11
on	1203	+	56
attacks	107	+	55.2
justice	62	+	52.61
soon	74	+	52.27
america	115	+	49.21
attack	167	+	49.2
just	248	+	48.34
keep	69	+	47.4
killed	324	+	46.34
families	75	+	45.1
captured	65	+	44.37
down	114	+	44.05
fighters	112	+	44.03
advice	72	+	43.81
message	88	+	43.67
we	768	+	42.7
under	184	+	39.74
never	180	+	39.64
all	613	+	38.98
join	69	+	38.82
airstrikes	68	+	38.62
know	211	+	38.02
free	85	+	36.87
again	77	+	36.03
are	1117	+	34.79
see	177	+	34.58
imam	129	+	33.16

children	143	+	33.09
libya	65	+	32.52
right	101	+	32.18
b	64	+	31.64
hope	62	+	31.18
country	67	+	30.52
syrian	120	+	30.23
h	82	+	29.17
must	116	+	28.73
kill	126	+	28.67
th	175	+	26.87
go	102	+	26.65
am	87	+	26.45
back	149	+	23.61
american	113	+	23.14
out	359	+	22.54
always	63	+	22.42
the	6762	-	3642.79
and	2741	-	2857.36
of	4385	-	1241.49
allah	490	-	1087.96
his	317	-	820.33
he	499	-	743.96
that	899	-	690.97
them	340	-	674.14
their	538	-	599.53
as	534	-	510.19
him	163	-	508.05
said	207	-	475.34
then	122	-	363.92
they	815	-	352.36
which	120	-	345.24
would	167	-	238.49
with	812	-	230.82
were	216	-	218.61
had	78	-	193.17
was	557	-	186.64
from	750	-	181.32
it	921	-	163.19
ibn	185	-	160.41
who	623	-	154.96
those	276	-	144.23
o	64	-	131.69
also	94	-	122.17

so	465	-	114.32
its	215	-	114.14
al	782	-	111.45
prophet	73	-	110.95
to	3954	-	106.89
not	778	-	105.52
religion	125	-	100.94
other	119	-	87.42
while	73	-	84.01
soldiers	113	-	82.54
for	1392	-	77.44
or	327	-	72.47
one	369	-	71.64
any	104	-	67.3
against	304	-	59.56
before	83	-	56.97
these	179	-	56.69
into	108	-	48.05
people	349	-	47.17
being	154	-	42.62
when	267	-	41.42
some	130	-	40.88
her	87	-	37.63
an	410	-	37.54
between	83	-	32.38
did	113	-	31.6
state	424	-	30.97
there	254	-	30.96
by	987	-	29.94
even	149	-	27.75
does	68	-	26.65
than	163	-	25.38
if	405	-	23.58
after	261	-	23.33
be	721	-	22.87
another	67	-	22.77
because	101	-	22.67
she	61	-	22.39
will	630	-	22.32

Anhang 16: Keyword List von Korpus C in Bezug zu Referenzkorpus A

Wort	Frequenz	Vorzeichen	Keyness
isis	50537	+	8438.15
you	81191	+	5807.19
your	36710	+	3396.22
i	42950	+	2943.13
quran	16085	+	2435.01
n	14225	+	2210.31
nn	12565	+	2157.43
me	20652	+	2083.83
t	19419	+	1878.04
is	84198	+	1658.52
ya	8081	+	1344.48
don	8119	+	1050.88
love	7831	+	991.48
allah	64634	+	944.15
side	6446	+	940.89
my	13831	+	937.69
can	11769	+	923.78
remember	6678	+	899.17
a	77324	+	836.69
forgive	5307	+	832.91
m	5893	+	817.57
coronavirus	4381	+	752.22
never	8485	+	745.47
trump	4252	+	730.07
muslim	13879	+	721.8
us	14456	+	715.18
have	25926	+	671.35
alhamdulillah	4033	+	656.12
good	9021	+	654.21
thankful	3851	+	641.84
caliphate	4868	+	610.94
make	8714	+	571.65
please	3698	+	567.65
nnin	3280	+	563.18
life	6106	+	562.56
someone	4634	+	550.45
tablighis	3138	+	538.8
oh	3172	+	535.99
arrested	3356	+	535.82
militancy	3102	+	530.45
barrier	3178	+	526.32
covid	3061	+	525.58
retweet	3049	+	523.52

seas	3088	+	521.57
maybe	3204	+	518.11
transgresses	3087	+	517.1
supporters	3880	+	516.57
india	3365	+	514.74
ameen	3023	+	508.26
armed	3685	+	500.57
service	3308	+	498.93
last	5377	+	488.54
afghanistan	4334	+	478.51
meeting	3344	+	473.41
yes	4258	+	472.79
need	4257	+	460.87
keep	3723	+	457.19
syria	4897	+	453.81
religious	3900	+	453.14
afghan	3153	+	448.71
pray	3156	+	447.27
pretty	2587	+	433.41
go	4719	+	425.74
terrorist	3073	+	425.45
pejorative	2416	+	414.83
harmless	2421	+	413.52
better	4320	+	413.06
released	3394	+	411.2
everything	3673	+	408.83
are	28876	+	399.79
u	3341	+	397.08
khorasan	2319	+	396.01
rss	2296	+	394.23
beautiful	2500	+	393.25
trust	3024	+	390.84
ll	2608	+	385.37
mother	3472	+	384.27
ungrateful	2307	+	381.08
bless	2522	+	374.68
happy	2331	+	374.67
learn	2640	+	373.13
today	4932	+	372.52
nbe	2128	+	365.38
biden	2114	+	362.98
time	7479	+	360.46
tomorrow	2342	+	360.04
re	3371	+	356.1
happiness	2250	+	352.48
islam	12117	+	347.07

health	2410	+	339.86
kabul	2199	+	339.65
ramadan	2742	+	334.5
nnthe	1938	+	332.76
neither	3235	+	332.4
history	3691	+	327.38
because	7604	+	316.92
terms	2484	+	316.59
attack	5069	+	314.97
swt	1818	+	312.15
want	3179	+	307.47
prayer	3890	+	306.17
world	5371	+	305.83
give	4660	+	304.92
things	3441	+	304.64
know	6287	+	303.94
sin	3136	+	303.74
every	6011	+	302.15
best	5208	+	300.86
led	3693	+	295.81
pakistan	2329	+	294.22
thing	2670	+	293.13
knows	3261	+	289.88
times	3657	+	288.92
surely	2872	+	288.28
pl	1684	+	286.98
civilians	1897	+	285.96
when	14129	+	285.03
feel	2142	+	284.9
dear	1936	+	284.62
our	10643	+	283.7
gallagher	1650	+	283.31
eddie	1645	+	282.45
posed	1730	+	275.79
islamic	17806	+	272.16
poor	1888	+	266.64
than	9553	+	266.58
shoot	1710	+	266.16
s	27230	+	260.49
no	11873	+	258.99
fighter	1851	+	256.49
falling	1627	+	251.95
how	6862	+	251.66
most	7005	+	250.25
responsibility	1987	+	247.78
establish	3219	+	246.18

grave	2018	+	245.79
stop	2517	+	245.56
am	3170	+	244.29
if	16807	+	238.5
ill	1636	+	237.43
burden	1623	+	237.18
holy	1598	+	236.89
always	2597	+	233.65
doesn	2125	+	230.7
mahdi	1658	+	227.59
abdullah	2418	+	226.67
injured	1734	+	223.56
really	1787	+	217.94
back	4546	+	216.26
get	2313	+	214.64
mercy	2787	+	214.03
chief	1577	+	213.87
claims	2200	+	210.57
process	1721	+	206.94
friend	1620	+	206.23
friends	1781	+	204.67
let	3620	+	204.26
greater	2474	+	204.25
iran	2591	+	198.22
parents	1832	+	197.8
death	3857	+	193.52
only	7697	+	193.17
soul	2167	+	193.15
help	2664	+	192.92
yourself	1936	+	191.9
understand	1876	+	190.23
about	6976	+	189.57
forces	3508	+	188.97
right	2981	+	185.31
new	3591	+	182.28
faith	2448	+	178.17
saw	2751	+	177.24
body	2056	+	175.84
look	2018	+	175.81
for	47330	+	175.19
just	5775	+	172.88
god	2514	+	167.02
food	1786	+	164.88
taliban	2091	+	159.95
true	2399	+	159.26
up	4794	+	158.09

recently	1631	+	155.68
sins	2037	+	154.56
grant	1936	+	154.52
nothing	3294	+	150.95
iraq	4712	+	148.83
peace	2415	+	143.08
why	2542	+	141.88
family	2822	+	138.25
down	2708	+	137.3
means	3428	+	136.76
all	13612	+	135.3
shall	2242	+	132.02
say	6280	+	128.66
says	2195	+	126.67
may	6429	+	122.63
on	25308	+	122.53
turkey	1982	+	120.76
obama	1553	+	118.69
still	1984	+	117.43
ve	1958	+	115
here	2724	+	113.37
base	1978	+	108.43
going	1737	+	104.39
doing	2138	+	104.36
country	1691	+	103.96
leader	2814	+	102.42
person	2799	+	101.57
ask	2663	+	100.19
like	5020	+	95.8
former	1920	+	91.88
home	1977	+	90.85
more	7533	+	88.36
put	1706	+	86.37
jannah	1804	+	85.27
light	1840	+	84.68
before	6125	+	84.04
days	2708	+	82.44
control	2352	+	81.8
name	2117	+	78.74
now	3046	+	78.74
blessing	1771	+	76.93
bring	1808	+	75.57
own	2942	+	73.34
but	11819	+	67.66
we	15528	+	64.44
think	1877	+	62.44

return	2256	+	61.88
cannot	1613	+	61.37
message	1584	+	58.84
law	1755	+	52.19
woman	1856	+	50.21
might	2722	+	48.19
praise	1964	+	45.49
has	12851	+	44.81
be	21939	+	43
state	14588	+	40.37
day	4735	+	38.82
at	11388	+	32.8
same	2439	+	29.49
th	3223	+	28.88
away	2308	+	28.85
between	4445	+	28.13
heart	1815	+	27.36
something	2016	+	25.18
bukhari	1642	+	24.53
the	155429	-	22099.97
and	78407	-	13068.49
their	8327	-	8075.75
of	96875	-	7379
al	6357	-	6121.56
said	4178	-	5218.08
his	11938	-	3860.2
they	15866	-	3111.31
as	13530	-	2968.59
would	2880	-	2870.77
its	2481	-	2376.24
were	4022	-	2268.26
that	28482	-	1888.41
upon	2420	-	1602.42
which	4811	-	1572.26
then	5140	-	1554.77
these	1923	-	1501.29
them	15751	-	1464.38
was	11402	-	1428.88
he	21899	-	1306.45
him	8878	-	1278.15
those	5749	-	1109.23
against	4780	-	938.97
from	17961	-	761.17
had	3195	-	738.72
or	6712	-	538.87
some	2149	-	525.72

by	18170	-	521.49
until	2099	-	503.95
while	1958	-	468.71
after	4492	-	359.54
prophet	2685	-	350.12
in	64693	-	334.72
fight	1715	-	308.05
killed	2291	-	295.98
many	2072	-	268.28
so	13072	-	203.65
among	2108	-	157.21
into	2941	-	156.5
muslims	6061	-	143.83
with	26632	-	132.57
also	4597	-	100.54
to	97373	-	85.55
being	4458	-	69.95
jihad	1755	-	68.63
any	3892	-	66.93
there	6680	-	49.07
over	3227	-	46.06
did	3355	-	43.9
she	1820	-	40.63
who	21090	-	37.44
her	3032	-	29.37
people	10084	-	27.96
should	2898	-	26.42

Anhang 17: Kritische Werte für den Chi-Quadrat-Test

ϕ	0,999	0,995	0,990	0,950	0,900	0,500	0,100	0,050	0,010	0,005	0,001
1	0,000	0,000	0,000	0,004	0,016	0,455	2,706	3,841	6,635	7,879	10,828
2	0,002	0,010	0,020	0,103	0,211	1,386	4,605	5,991	9,210	10,597	13,816
3	0,024	0,072	0,115	0,352	0,584	2,366	6,251	7,815	11,345	12,838	16,266
4	0,091	0,207	0,297	0,711	1,064	3,357	7,779	9,488	13,277	14,860	18,467
5	0,210	0,412	0,554	1,145	1,610	4,351	9,236	11,070	15,086	16,750	20,515
6	0,381	0,676	0,872	1,635	2,204	5,348	10,645	12,592	16,812	18,548	22,458
7	0,598	0,989	1,239	2,167	2,833	6,346	12,017	14,067	18,475	20,278	24,322
8	0,857	1,344	1,646	2,733	3,490	7,344	13,362	15,507	20,090	21,955	26,124
9	1,152	1,735	2,088	3,325	4,168	8,343	14,684	16,919	21,666	23,589	27,877
10	1,479	2,156	2,558	3,940	4,865	9,342	15,987	18,307	23,209	25,188	29,588
11	1,834	2,603	3,053	4,575	5,578	10,341	17,275	19,675	24,725	26,757	31,264
12	2,214	3,074	3,571	5,226	6,304	11,340	18,549	21,026	26,217	28,300	32,909
13	2,617	3,565	4,107	5,892	7,042	12,340	19,812	22,362	27,688	29,819	34,528
14	3,041	4,075	4,660	6,571	7,790	13,339	21,064	23,685	29,141	31,319	36,123
15	3,483	4,601	5,229	7,261	8,547	14,339	22,307	24,996	30,578	32,801	37,697

Anhang 18: Berechnung der Chi-Quadrat-Prüfwerte verschiedener Terme

Für „allah“ in Korpus B: siehe Kapitel 3.5

Für „allah“ in Korpus C:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort x	7674	64634	72308
4	Alle anderen Wörter	700878	4956943	5657821
5	Total	708552	5021577	5730129
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort x	8941,156127	63366,8439	72308
10	Alle anderen Wörter	699610,8439	4958210,16	5657821
11	Total	708552	5021577	5730129
12				
13				
14	Chi Quadrat	207,542023	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „state“ in Korpus B:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort X	2164	424	2588
4	Alle anderen Wörter	706388	185945	892333
5	Total	708552	186369	894921
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort X	2049,044079	538,955921	2588
10	Alle anderen Wörter	706502,9559	185830,044	892333
11	Total	708552	186369	894921
12				
13				
14	Chi Quadrat	31,0584774	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „state“ in Korpus C:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort x	2164	14588	16752
4	Alle anderen Wörter	706388	5006989	5713377
5	Total	708552	5021577	5730129
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort x	2071,447799	14680,5522	16752
10	Alle anderen Wörter	706480,5522	5006896,45	5713377
11	Total	708552	5021577	5730129
12				
13				
14	Chi Quadrat	4,73255097	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „prophet“ in Korpus B:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort x	927	73	1000
4	Alle anderen Wörter	707625	186296	893921
5	Total	708552	186369	894921
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort x	791,7480984	208,251902	1000
10	Alle anderen Wörter	707760,2519	186160,748	893921
11	Total	708552	186369	894921
12				
13				
14	Chi Quadrat	111,069883	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „prophet“ in Korpus C:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort X	927	2685	3612
4	Alle anderen Wörter	707625	5018892	5726517
5	Total	708552	5021577	5730129
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort X	446,6373836	3165,36262	3612
10	Alle anderen Wörter	708105,3626	5018411,64	5726517
11	Total	708552	5021577	5730129
12				
13				
14	Chi Quadrat	589,904148	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „jihad“ in Korpus B:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort X	639	1916	2555
4	Alle anderen Wörter	707913	184453	892366
5	Total	708552	186369	894921
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort X	2022,916392	532,083608	2555
10	Alle anderen Wörter	706529,0836	185836,916	892366
11	Total	708552	186369	894921
12				
13				
14	Chi Quadrat	4559,26129	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

Für „jihad“ in Korpus C:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort x	639	1755	2394
4	Alle anderen Wörter	707913	5019822	5727735
5	Total	708552	5021577	5730129
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort x	296,0271031	2097,9729	2394
10	Alle anderen Wörter	708255,9729	5019479,03	5727735
11	Total	708552	5021577	5730129
12				
13				
14	Chi Quadrat	453,621765	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

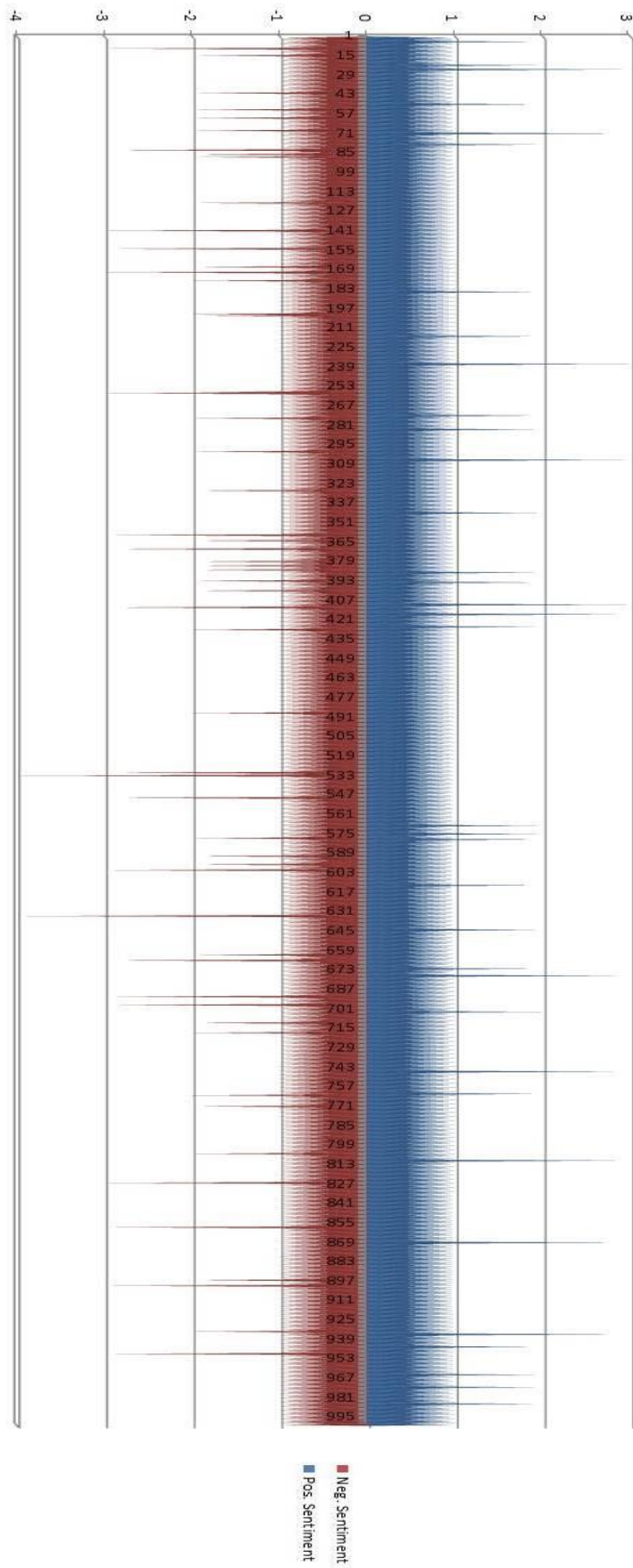
Für „islam“ in Korpus B:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort X	1190	1377	2567
4	Alle anderen Wörter	707362	184992	892354
5	Total	708552	186369	894921
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort X	2032,417369	534,582631	2567
10	Alle anderen Wörter	706519,5826	185834,417	892354
11	Total	708552	186369	894921
12				
13				
14	Chi Quadrat	1681,51318	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

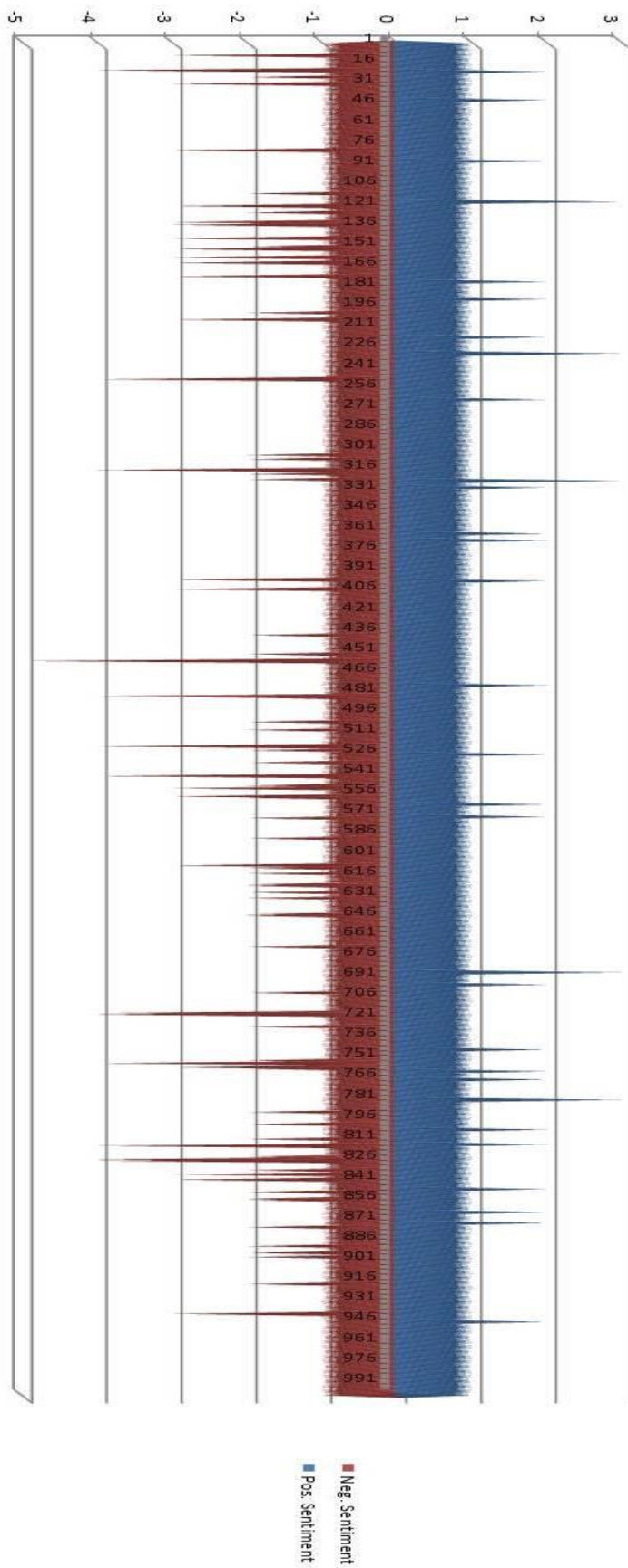
Für „islam“ in Korpus C:

	A	B	C	D
1	Beobachtete Werte			
2		Korpus A	Korpus B	Total
3	Frequenz von Wort x	1190	12117	13307
4	Alle anderen Wörter	707362	5020200	5727562
5	Total	708552	5032317	5740869
6				
7	Erwartete Werte			
8		Korpus A	Korpus B	Total
9	Frequenz von Wort x	1642,382271	11664,6177	13307
10	Alle anderen Wörter	706909,6177	5020652,38	5727562
11	Total	708552	5032317	5740869
12				
13				
14	Chi Quadrat	142,480158	signifikant	
15	df	1		
16				
17				
18	p	Kritische Werte		
19	0,05	3,841458821		
20	0,01	6,634896601		
21	0,001	10,82756617		
22	0,0001	15,13670523		

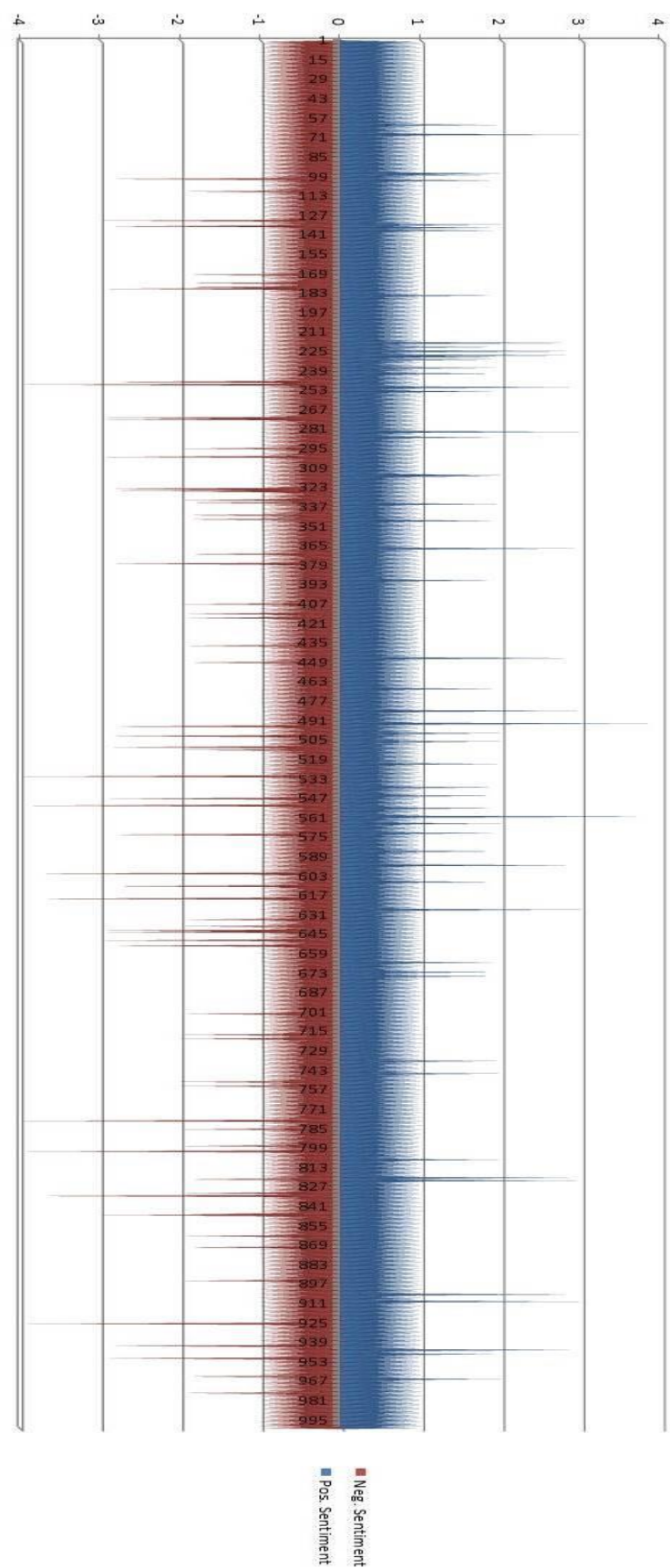
Anhang 19: Sentiment-Score der Top-1000 Terme in Korpus A



Anhang 20: Sentiment-Score der Top-1000 Terme in Korpus B



Anhang 21: Sentiment-Score der Top-1000 Terme in Korpus C



Anhang 22: Ausgewertete Propaganda-Magazine

- Dabiq* (2014a). The Return of Khilafah. *Dabiq Issue 1*.
- Dabiq* (2014b). The Flood. *Dabiq Issue 2*.
- Dabiq* (2014c). A Call to Hijrah. *Dabiq Issue 3*.
- Dabiq* (2014d). The Failed Crusade. *Dabiq Issue 4*.
- Dabiq* (2014e). Remaining and Expanding. *Dabiq Issue 5*.
- Dabiq* (2014f). Al-Qa'idah of Waziristan: A Testimony From Within. *Dabiq Issue 6*.
- Dabiq* (2015a). From Hypocrisy to Apostasy. *Dabiq Issue 7*.
- Dabiq* (2015b). Shari'ah Alone Will Rule Africa. *Dabiq Issue 8*.
- Dabiq* (2015c). They Plot and Allah Plots. *Dabiq Issue 9*.
- Dabiq* (2015d). The Laws of All or the Laws of Men. *Dabiq Issue 10*.
- Dabiq* (2015e). From the Battle of Al-Ahzab to the War of Coalitions. *Dabiq Issue 11*.
- Dabiq* (2015f). Just Terror. *Dabiq Issue 12*.
- Dabiq* (2016a). The Rafidah: From Ibn Sa'ba to the Dajjal. *Dabiq Issue 13*.
- Dabiq* (2016b). The Murtadd Brotherhood. *Dabiq Issue 14*.
- Dabiq* (2016c). Breaking the Cross. *Dabiq Issue 15*.
- Rumiyah* (2016a). *Rumiyah Issue 1*.
- Rumiyah* (2016b). *Rumiyah Issue 2*.
- Rumiyah* (2016c). *Rumiyah Issue 3*.
- Rumiyah* (2016d). *Rumiyah Issue 4*.
- Rumiyah* (2017a). *Rumiyah Issue 5*.
- Rumiyah* (2017b). *Rumiyah Issue 6*.
- Rumiyah* (2017c). *Rumiyah Issue 7*.
- Rumiyah* (2017d). *Rumiyah Issue 8*.
- Rumiyah* (2017e). *Rumiyah Issue 9*.
- Rumiyah* (2017f). *Rumiyah Issue 10*.
- Rumiyah* (2017g). *Rumiyah Issue 11*.
- Rumiyah* (2017h). *Rumiyah Issue 12*.
- Rumiyah* (2017i). *Rumiyah Issue 13*.